

AD-A065 650

OHIO STATE UNIV RESEARCH FOUNDATION COLUMBUS
USAF-ASEE (1978) SUMMER FACULTY RESEARCH PROGRAM (WPAFB). VOLUM--ETC(U)
NOV 78 C D BAILEY

F44620-76-C-0052

AFOSR-TR-79-0231

NL

UNCLASSIFIED

1 OF 6

AD
A065650



AFOSR-TR- 79-0281

LEVEL III

Q
NW

AD A0 65650

1978

A065651

USAF-ASEE
SUMMER FACULTY
RESEARCH PROGRAM
WPAFB

DDC
RECEIVED
MAR 14 1979
C

DDC FILE COPY

Volume I of II
Research Report

OSU

The Ohio State University
Research Foundation
Columbus, Ohio 43212

November 1978

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFOSR)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release in accordance with AFR 190-12 (7b).
Distribution is unlimited.
A. D. SLOSE
Technical Information Office

79 03 12 012

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR-79-0231	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) USAF-ASEE (1978) Summer Faculty Research Program (Volume I).	5. TYPE OF REPORT & PERIOD COVERED Final rept. 1 Jan 76 - 30 Sep 78.	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Professor Cecil D. Bailey	8. CONTRACT OR GRANT NUMBER(s) F44620-76-C-0052	9. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F, 2307/D1
10. PERFORMING ORGANIZATION NAME AND ADDRESS Ohio State University Research Foundation Columbus OH 43212	11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/XOP Bldg. 410 Bolling AFB DC 20332	12. REPORT DATE November 1978
13. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 525p.	14. NUMBER OF PAGES 266	15. SECURITY CLASS. (of this report) UNCLASSIFIED
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The fourth USAF-ASEE Summer Faculty Program was conducted at Wright-Patterson AFB during the summer of 1978. A total of twenty-five research associates (participants) from sixteen universities in eleven states participated in the program. The program was again highly successful as judged from the participants' written reports of research accomplished, verbal contacts with those in the program (both participants and associated Air Force personnel) and upon written evaluations of the program by both the		

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

participants and Air Force personnel.

07A

TABLE OF CONTENTS

Volume I

AF WRIGHT AERONAUTICAL LABORATORIES

Partial contents

AIR FORCE FLIGHT DYNAMICS LABORATORY

- Regularization of Ill-Posed Problems; Charles W. Groetsch
- Unconditionally Stable Second Order
Accurate Method for Transonic Flutter
Calculations; Czeslaw P. Kentzer
- Large Amplitude Response of Complex
Structures Due to High Intensity Noise; . . Chuh Mei
- Unified Synthesis of Manual and Auto-
matic Control Applied to Integrated Fire
and Flight Control; David K. Schmidt

AIR FORCE AERO-PROPULSION LABORATORY

- Catalytic Flame Stabilization for Air-
craft Afterburners; Donald R. Jenkins
- Investigation of Cadmium Deposition
Reactions on Cadmium and Nickel Elec-
trodes by Cyclic Voltammetry; Yuen-Koh Kao
- Turbulent Flow Measurements for Sudden
Expansion Cylindrical Tube by Using
Laser Doppler Velocimeter; (LDV) Chris C. Lu
- Preliminary Design Procedure for High
Power Density MHD Generators; Pau-Chang Lu

AIR FORCE MATERIALS LABORATORY

- Economic Modeling and Computer-Aided
Process Planning for Sheet Metal
Operations; Brian K. Lambert
- Feasibility Study on the Use of the ^{Using}
EPR Technique on Detector Grade Silicon; . . . George K. Miner
- A Simulation Metamodel; Robert E. Young

AIR FORCE AVIONICS LABORATORY

- Analysis of the GPS Receiver Loss-of-
Lock Problem; William L. Brogan
- and Physics of Matrix Cathodes Thomas P. Graham

79 03 12 012

AF AVIONICS LAB., Continued

- Efficient Fault Analysis in Analog
Circuits Alfred T. Johnson, Jr.
- Electrical Properties of Magnesium and
Germanium Implanted Gallium Arsenide Frank L. Pedrotti

Volume II

AEROSPACE MEDICAL RESEARCH LABORATORY

- Mathematical Modeling of the Response of
the Vascular System to Time-Dependent
Accelerations Xavier J.R. Avula
- Evoked Response Measures of Resource Allo-
cation: Effects of Varying the Primary
Task Workload Lloyd F. Elfner
- An Information-Theoretic Description of
the CxC System Duane G. Leet
- The Effects of Bromchlorofluoromethane
(1211) on Canine Purkinje Fibers S. Mark Strauch
- Steady State Evoked Responses: Interaction
with Cognitive Task Load Glenn F. Wilson

AIR FORCE HUMAN RESOURCES LABORATORY

- Texture Modeling and Generation Using
Partial Difference Equations Richard J. Bethke
- Complexity Judgments of Computer Gen-
erated Simulations of Electro-Optical
Displays Moira K. LeMay
- A Model of Performance Effectiveness in the
Air Force Maintenance System Hewitt H. Young

AIR FORCE LOGISTICS COMMAND

- An Investigation of Production Leadtime
Forecasting for Air Logistics Centers Michael J. Cleary
- An Experimental Design for Selecting a
Forecasting Model for Predicting LRU
Failure Rates under Change Gordon K. Constable

AIR FORCE FLIGHT DYNAMICS LABORATORY

Research Associates:

Charles W. Groetsch, University of Cincinnati

Czeslaw P. Kentzer, Purdue University

Chuh Mei, University of Missouri -- Rolla

David K. Schmidt, Purdue University

AFFDL-TR-78-

REGULARIZATION OF ILL-POSED PROBLEMS

C.W. Groetsch

Regularization of Ill-Posed Problems

C. W. Groetsch

Department of Mathematical Sciences

University of Cincinnati

Cincinnati, Ohio 45221

ABSTRACT

Some examples of linear ill-posed problems in engineering are given and a general class of regularization methods for ill-posed linear operator equations is studied. Rates of convergence for the general method are established under various assumptions on the data. Applications are given to a number of iterative and noniterative regularization algorithms.

ACQUISITION FOR	
NTIS	Whole Section <input checked="" type="checkbox"/>
DDC	Full Section <input type="checkbox"/>
UNCLASSIFIED	<input type="checkbox"/>
RECLASSIFIED	
BY	
DATE	
L	
A	

FOREWORD

This report describes work performed in the Air Force Flight Dynamics Laboratory during the summer of 1978. The research was supported by the Air Force Office of Scientific Research through the USAF-ASEE Summer Faculty Research Program (WPAFB), Contract F44620-76-C-0052, The Ohio State University Research Foundation, Columbus, Ohio.

TABLE OF CONTENTS

<u>Section</u>	<u>Title</u>	<u>Page</u>
I	Introduction	1
II	Generalized Inverses	5
III	A General Method	9
IV	Specific Methods	17
	References	21

SECTION I

INTRODUCTION

The concept of a well-posed problem was formulated by Hadamard early in this century. In broad terms, a problem is well-posed in the sense of Hadamard if it has a unique solution which depends continuously on the data of the problem. Specifically, if T is a transformation from a metric space X into a metric space Y , then the problem

$$Tx = b \tag{1}$$

is said to be well-posed if

- (i) for each $b \in Y$ there is a solution $x \in X$,
- (ii) the solution x is unique, and
- (iii) the solution x depends continuously on the "data" b .

A problem which is not well-posed is called "ill-posed." Ill-posed problems have been intensively studied during the last fifteen years, especially by Soviet mathematicians (see [12],[23]), because of their importance in many engineering applications (see [11] and [14] for specific areas of Air Force interest). In this report we will be concerned with linear ill-posed problems, that is, we will study the problem (1) where T is a linear operator on Hilbert space. A typical problem of this type is the integral equation of the first kind

$$\int_a^d k(s,t)x(s)ds = b(t) \quad (2)$$

where the kernel k is a member of $L^2([a,d] \times [a,d])$ (the space of Lebesgue square integrable functions on the rectangle $[a,d] \times [a,d]$) and $b \in L^2[a,d]$ (we allow a or d to be infinite). Such equations are notoriously ill-posed. For example, if $k(s,t) = t + c$, then (2) can have a solution only if b is a linear function, violating (i). If $k(s,t) = \sin(s)$ and $b(t) = 2$, then by the well-known orthogonality relations,

$$\int_0^\pi k(s,t)(1 + \sin(ms))ds = b(t), \quad m = 2, 3, \dots$$

which violates (ii). Far more serious is the fact that (iii) is violated for equations of type (2). Indeed, by the Riemann-Lebesgue lemma, for arbitrary A ,

$$\int_0^1 k(s,t) A \sin(m\pi s)ds \rightarrow 0 \quad \text{as } m \rightarrow \infty,$$

and hence solutions do not depend continuously on the data.

Numerical methods for analyzing ill-posed linear problems are particularly important because a large number of engineering problems have the form (2). Consider for example the one dimensional heat equation

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t}, \quad u(x,0) = h(x).$$

It is well known that the temperature distribution $f(x) = u(x,T)$ at some time $T > 0$ can be expressed in terms of the initial temperature distribution $h(x)$ by

$$f(x) = \frac{1}{2\sqrt{\pi T}} \int_{-\infty}^{\infty} \exp(-(x-\tau)^2/(4T)) h(\tau) d\tau.$$

The "inverse" problem of determining the initial temperature distribution $h(x)$, given the distribution $f(x)$ at the later time, is of considerable interest and is an ill-posed problem of type (2).

Another problem of type (2) is the numerical differentiation problem. The n th derivative of a given function $b(t)$ (with $b(0) = b'(0) = \dots = b^{(n-1)}(0) = 0$) satisfies

$$\int_0^t \frac{1}{(n-1)!} (t-s)^{n-1} x(s) ds = b(t).$$

This problem has been studied extensively within the context of ill-posed problems by Cullum [3], Franklin [5] and others.

Another example is afforded by the work of Lee [13] and Provencher [18] on the determination of the molecular weight distribution of a solute from centrifuge data. In this example the molecular weight distribution $f(m)$ satisfies

$$U(x) = \int_0^{\infty} \frac{\lambda^2 m^2 e^{-\lambda m x}}{1 - e^{-\lambda m}} f(m) dm$$

where U is a function which is proportional to the measured concentration gradient and λ is a constant which is proportional to the square root of the rotor speed.

As a final example, we give the two dimensional integral equation

$$\int_{\Omega} \frac{p(x',y') dx' dy'}{[(x-x')^2 + (y-y')^2]^{\frac{1}{2}}} = \delta - f_1(x,y) - f_2(x,y)$$

which was studied by Singh and Paul [21] and concerns the pressure distribution in the contact of nonconforming elastic bodies.

Integral equations of the first kind also arise in the determination of the shape of conducting bodies from backscattered electromagnetic radiation ([16],[17]), seismic prospecting [2], antenna theory [4], remote probing of the atmosphere ([22],[24]), medical tomography [6] and system identification ([1],[15]).

SECTION II

GENERALIZED INVERSES

We will henceforth assume that H_1 and H_2 are Hilbert spaces and that $T: H_1 \rightarrow H_2$ is a bounded linear operator. The inner product and norm in each space will be denoted by (\cdot, \cdot) and $||\cdot||$, respectively. The range and nullspace of T will be denoted by $R(T)$ and $N(T)$, respectively. Our task is to solve the ill-posed problem (1) for $x \in H_1$ given $b \in H_2$. Of course, if $b \notin R(T)$ then (1) is violated and there is no solution. In such a case we might reasonably adopt the more flexible attitude of replacing b in the right hand side of (1) by the point in $R(T)$ which is nearest to b . However, if $R(T)$ is not closed, such a closest point may not exist. We are then led to accept as a generalized solution any vector $u \in H_1$ which satisfies

$$Tu = Pb \quad (3)$$

where P is the projection of H_2 onto $\overline{R(T)}$, the closure of $R(T)$. Any vector u satisfying (3) is called a least squares solution of equation (1). We note that a least squares solution will exist for any vector b whose projection onto $\overline{R(T)}$ lies in $R(T)$, i.e., for all vectors b in the dense subspace $R(T) \oplus R(T)^\perp$ of H_2 . It is not difficult to see that least squares solutions may also be characterized as vectors $u \in H_1$ which satisfy either of the conditions

$$||Tu - b|| \leq ||Tx - b||, \text{ for all } x \in H_1, \quad (4)$$

or

$$T^*Tu = T^*b \quad (5)$$

where T^* is the adjoint of T (see [7] for a proof of this and other simple facts pertaining to this section).

We have seen that if we consider least squares solutions instead of traditional solutions, then difficulty (i) is to a certain extent obviated. The problem of nonuniqueness, however, remains at this point. Indeed, if $N(T) \neq \{0\}$ then there may be infinitely many least squares solutions, for if u is a least squares solution, then so is $u+v$ for any $v \in N(T)$. Fortunately, there is a natural way of selecting a least square solution which is unique in a certain sense. We see from (5) that the set of all least squares solutions is a closed convex set. This set therefore contains a unique vector of smallest norm and it is this vector which we will accept as the unique generalized solution of equation (1). Let $\mathcal{D}(T^+) = R(T) \oplus R(T)^\perp$. The operator

$$T^+ : \mathcal{D}(T^+) \rightarrow H_1$$

which associates with each $b \in \mathcal{D}(T^+)$ the unique least squares solution of equation (1) with minimal norm is called the generalized inverse of T . It is not difficult to show that T^+ is a closed linear operator (see [7]). If T^+ were continuous then problems (i), (ii), (iii) would be solved, at least for $b \in \mathcal{D}(T^+)$. But alas this is not the case. It is not difficult to show that T^+ is continuous if and only if $R(T)$ is closed. Unfortunately the range of an integral operator is closed if and only if its kernel is degenerate (see [7]). We are therefore led to seek approximations to T^+ by bounded linear

operators. Such approximations, when applied to b , are called regularizers of equation (1).

SECTION III

A GENERAL METHOD

We will denote the operator T^*T by \tilde{T} and the operator TT^* by \hat{T} . Note that \tilde{T} and \hat{T} are self adjoint linear operators whose spectra lie in the interval $[0, ||T||^2]$. If $0 \notin \sigma(\tilde{T})$ (the spectrum of \tilde{T}), then by (5) we have $T^\dagger = \tilde{T}^{-1}T^*$. In general, however, $0 \in \sigma(\tilde{T})$, but this last equation nevertheless leads us to seek approximations to T^\dagger by operators of the form $U(\tilde{T})T^*$ where U is a continuous function on $[0, ||T||^2]$ which approximates the function $f(t) = t^{-1}$ in some sense. Specifically, we will consider a family (net) of real valued functions $\{U_\beta(t) : \beta \in S\}$, indexed by a subset S of the positive real numbers with $\infty \in \bar{S}$, where each U_β is continuous on $[0, ||T||^2]$ and such that

$$|tU_\beta(t)| \leq M \quad \text{for all } t \text{ and } \beta \quad (6)$$

and

$$U_\beta(t) \rightarrow t^{-1} \quad \text{as } \beta \rightarrow \infty \text{ for each } t \neq 0. \quad (7)$$

The following is proved in [7].

Theorem 1. Suppose $b \in \mathcal{N}(T^\dagger)$ and let $x_\beta = U_\beta(\tilde{T})T^*b$. Then $x_\beta \rightarrow T^\dagger b$ as $\beta \rightarrow \infty$.

To this we now add,

Theorem 2. If $b \notin \mathcal{N}(T^\dagger)$, then $\{x_\beta\}$ has no weakly convergent subnet.

Proof. Suppose $\{x_{\beta_\gamma}\}$ is a subnet of $\{x_\beta\}$ which converges weakly to $z \in H_1$, denoted $x_{\beta_\gamma} \xrightarrow{w} z$. By the weak continuity of bounded linear operators we then have $Tx_{\beta_\gamma} \xrightarrow{w} Tz$.

Now,

$$\begin{aligned} P_b - T x_\beta &= P_b - T U_\beta(\tilde{T}) T^* b \\ &= P_b - \hat{T} U_\beta(\hat{T}) P_b. \end{aligned}$$

However, by (6) and (7), the operator $\hat{T} U_\beta(\hat{T})$ converges pointwise to the projection of H_2 onto $N(\hat{T})^\perp = N(T^*)^\perp = \overline{R(T)}$. Therefore $P_b - T x_\beta \rightarrow 0$. It then follows that $P_b = T z$, a contradiction. #

In the proof above we have used the fact that $U_\beta(\tilde{T}) T^* = T^* U_\beta(\hat{T})$. This is easy to see if U_β is a polynomial. In the general case the identity follows from the Weierstrass approximation theorem. Using the fact that bounded sets in Hilbert space are weakly compact, we have:

Corollary 3. If $b \notin \mathcal{D}(T^+)$, then $\|x_\beta\| \rightarrow \infty$ as $\beta \rightarrow \infty$.

Theorem 1 and Corollary 3 demonstrate dramatically the unequivocal nature of the approximations $\{x_\beta\}$.

Several authors have established rates of convergence for various approximations to $T^+ b$ under the stronger assumption that $P_b \in R(\hat{T})$ (see [20],[9],[10]). We see from Corollary 3 that the very least we must require to get convergence at all is that $P_b \in R(T)$. In order to strengthen this condition only slightly and thereby obtain a rate of convergence we note that

$$R(T) = R(T P_{N(T)^\perp})$$

and, in the pointwise sense,

$$P_{N(T)}^\perp = \lim_{\nu \rightarrow 0^+} \tilde{T}^\nu.$$

It therefore seems reasonable to replace the hypothesis $b \in \mathcal{N}(T^\dagger)$, i.e., $Pb \in R(T)$, by the hypothesis $Pb \in R(\tilde{T}^\nu)$ for some $\nu > 0$. In order to gauge the rate of convergence we will replace (7) by the stronger condition

$$t^\nu |1 - tU_\beta(t)| \leq \omega(\beta, \nu) \quad \text{for } \nu > 0 \quad (8)$$

where $\omega(\beta, \nu) \rightarrow 0$ as $\beta \rightarrow \infty$ for each $\nu > 0$ (the case $\nu = 1$ was considered in [8]).

Lemma 4. If $\nu > 0$, then $R(\tilde{T}^\nu) \subseteq N(T)^\perp$.

Proof. Suppose $\{f_n\}$ is a sequence of continuous real valued functions on $[0, \|T\|^2]$ such that $f_n(t) \rightarrow t^{\nu-1}$ for $t \neq 0$ and $tf_n(t)$ is uniformly bounded (for example, we may take $f_n(t) = t^{\nu-1}$ for $t \geq 1/n$ and $f_n(t) = n^{2-\nu}t$ for $0 \leq t \leq 1/n$). Let $\{E_t\}$ be the resolution of the identity generated by the self-adjoint operator \tilde{T} . By the bounded convergence theorem we then have

$$\begin{aligned} \tilde{T}^\nu y &= \int_0^{\|T\|^2} t^\nu dE_t y = \int_0^{\|T\|^2} \lim_n tf_n(t) dE_t y \\ &= \lim_n \int_0^{\|T\|^2} tf_n(t) dE_t y = \lim_n \tilde{T}f_n(\tilde{T})y \in N(T)^\perp. \# \end{aligned}$$

We now state a rate of convergence result. The vector $T^\dagger b$ will be denoted by x and the error $x - x_\beta$ by e_β .

Theorem 5. If $Pb = T\tilde{T}^v w$, where $v > 0$, then $\|e_\beta\| \leq \omega(\beta, v) \|w\|$.

Proof. Since $Tx = Pb = T\tilde{T}^v w$ and since $x - \tilde{T}^v w \in N(T)^\perp$, we see that $x = \tilde{T}^v w$. Now,

$$\begin{aligned} x_\beta &= U_\beta(\tilde{T})T^*b = U_\beta(\tilde{T})T^*Pb \\ &= U_\beta(\tilde{T})\tilde{T}x = U_\beta(\tilde{T})\tilde{T}^{v+1}w. \end{aligned}$$

Therefore $e_\beta = x - x_\beta = \tilde{T}^v(I - U_\beta(\tilde{T})\tilde{T})w$.

By the Spectral Mapping Theorem and Radius Formula, we then have

$$\|e_\beta\| \leq \omega(\beta, v) \|w\|. \#$$

In our next result we become more cavalier in our assumptions on the data.

Lemma 6. If $Pb = \hat{T}^v w$ where $v \geq 1$, then $\|e_\beta\|^2 \leq \omega(\beta, v-1) \|Te_\beta\| \|w\|$.

Proof. As in the previous proof we find that $x = T^*\hat{T}^{v-1}w$. Also,

$$\begin{aligned} x_\beta &= U_\beta(\tilde{T})T^*Pb = U_\beta(\tilde{T})T^*\hat{T}^v w \\ &= T^*U_\beta(\hat{T})\hat{T}^v w. \end{aligned}$$

Therefore $e_\beta = x - x_\beta = T^*(I - U_\beta(\hat{T})\hat{T})\hat{T}^{v-1}w$, and

$$\begin{aligned} \|e_\beta\|^2 &= (e_\beta, T^*(I - U_\beta(\hat{T})\hat{T})\hat{T}^{v-1}w) \\ &= (Te_\beta, (I - U_\beta(\hat{T})\hat{T})\hat{T}^{v-1}w) \leq \omega(\beta, v-1) \|w\| \|Te_\beta\|. \# \end{aligned}$$

Theorem 7. If $Pb = \hat{T}^v w$ where $v \geq 1$, then $\|e_\beta\|^2 \leq \omega(\beta, v)\omega(\beta, v-1) \|w\|$.

Proof. In Lemma 6 we saw that

$$e_\beta = T^*(I - U_\beta(\hat{T})\hat{T})\hat{T}^{v-1}w,$$

therefore

$$\tilde{Te}_\beta = T^* T^v (I - U_\beta(\hat{T})\hat{T})w.$$

We then have

$$\begin{aligned} ||Te_\beta||^2 &= (\tilde{Te}_\beta, e_\beta) = (T^v(I - U_\beta(\hat{T})\hat{T})w, Te_\beta) \\ &\leq \omega(\beta, v) ||Te_\beta||, \text{ i.e., } ||Te_\beta|| \leq \omega(\beta, v). \end{aligned}$$

Substituting into the result of Lemma 6 completes the proof. #

In the next section we will give a number of examples of specific computational techniques to which the above results apply.

We have avoided for long enough the problem of polluted data. We now take up this question. Suppose that the data b is the result of measurements so that instead of b we have in our possession a corrupted version b^ϵ satisfying $||b - b^\epsilon|| \leq \epsilon$. We operate on the vector b^ϵ to obtain the approximations x_β^ϵ given by

$$x_\beta^\epsilon = U_\beta(\tilde{T})T^*b^\epsilon.$$

Let $\phi(\beta) = \sup\{|tU_\beta(t)| : t \in [0, ||T||^2]\}$, and recall that $\phi(\beta)$ is bounded (by (6)).

Lemma 8. $||Tx_\beta - Tx_\beta^\epsilon|| \leq \epsilon \phi(\beta).$

Proof. $\tilde{T}(x_\beta - x_\beta^\epsilon) = \tilde{T}U_\beta(\tilde{T})T^*(b - b^\epsilon)$, therefore

$$\begin{aligned} ||Tx_\beta - Tx_\beta^\epsilon||^2 &= (\tilde{T}(x_\beta - x_\beta^\epsilon), x_\beta - x_\beta^\epsilon) \\ &= (\tilde{T}U_\beta(\tilde{T})T^*(b - b^\epsilon), x_\beta - x_\beta^\epsilon) \\ &= (\hat{T}U_\beta(\hat{T})(b - b^\epsilon), T(x_\beta - x_\beta^\epsilon)) \end{aligned}$$

$$\leq \phi(\beta) \|b - b^\epsilon\| \|T(x_\beta - x_\beta^\epsilon)\|$$

$$\leq \epsilon \phi(\beta) \|Tx_\beta - Tx_\beta^\epsilon\|. \#$$

Suppose now that $g(\beta) = \sup\{|U_\beta(t)| : t \in [0, \|T\|^2]\}$. We note that

$$g(\beta) \rightarrow \infty \text{ as } \beta \rightarrow \infty. \quad (9)$$

Indeed, if this were not the case, then there would be a constant L such that $|U_\beta(t)| < L$ for all t and β . But then $|tU_\beta(t)| \leq Lt \rightarrow 0$ as $t \rightarrow 0$, contradicting (7).

Lemma 9. $\|x_\beta - x_\beta^\epsilon\| \leq \epsilon \sqrt{g(\beta)\phi(\beta)}.$

Proof. Since $x_\beta - x_\beta^\epsilon = T^*U_\beta(\hat{T})(b - b^\epsilon)$, we have, by use of Lemma 8,

$$\begin{aligned} \|x_\beta - x_\beta^\epsilon\|^2 &= (x_\beta - x_\beta^\epsilon, T^*U_\beta(\hat{T})(b - b^\epsilon)) \\ &= (T(x_\beta - x_\beta^\epsilon), U_\beta(\hat{T})(b - b^\epsilon)) \\ &\leq \epsilon^2 \phi(\beta) g(\beta). \# \end{aligned}$$

Suppose now that $Pb = \hat{T}w$ (we could also use the other hypotheses considered above, but we choose to consider this simple case to illustrate the ideas). By the triangle inequality we have

$$\|x - x_\beta^\epsilon\| \leq \|x - x_\beta\| + \|x_\beta - x_\beta^\epsilon\|.$$

Lemma 9 and Theorem 7, then give

Theorem 10. If $Pb = \hat{T}w$, then

$$||x - x_{\beta}^{\epsilon}|| \leq (||w||\omega(\beta,1)\omega(\beta,0))^{\frac{1}{2}} + \epsilon(g(\beta)\phi(\beta))^{\frac{1}{2}}.$$

The first term on the right hand side of this inequality goes to zero as $\beta \rightarrow \infty$. However, by (9) and (7), the second term becomes infinitely large as $\beta \rightarrow \infty$. This illustrates the classic dilemma in the numerical treatment of ill-posed problems. Even if computations are performed exactly, small errors in the data may eventually grow and overpower the approximations.

In view of Theorem 10, the question naturally arises as to whether it is ever possible to obtain convergent approximations even if the data can be measured as precisely as desired. Specifically, is there an effective way of choosing a "stopping parameter" $\beta(\epsilon)$ such that $e_{\beta(\epsilon)} \rightarrow 0$ as $\epsilon \rightarrow 0$? This problem of choice of regularization parameters is of great import and still has not been satisfactorily answered. For the wide class of methods considered here the question is particularly difficult, for as we shall see in the next section, the parameter may take on discrete or continuous values depending upon the specific method under consideration.

SECTION IV

SPECIFIC METHODS

In this section we will consider some specific choices for the functions $\{U_\beta(t)\}$ and we will find functions $\omega(\beta, \nu)$ which determine rates of convergence. The index set S in all examples below will be either the set of nonnegative reals or nonnegative integers. In the discrete case, the parameter β will be denoted by n .

As a first example we consider Showalter's integral formula [19]:

$$T^+ b = \int_0^\infty \exp(-u\tilde{T}) T^* b du.$$

The functions U_β for this example have the form

$$U_\beta(t) = \int_0^\beta \exp(-ut) du$$

and may be motivated in terms of Borel summability [7]. It is not difficult to see that a function $\omega(\beta, \nu)$ satisfying (8) is given by

$$\omega(\beta, \nu) = \beta^{-\nu} \quad (\nu > 0).$$

The choice $U_\beta(t) = (t + \beta^{-1})^{-1}$ ($\beta > 0$) leads to Tychonov's regularization of order zero [23]. Here one can readily verify that

$$\omega(\beta, \nu) = \beta^{-\nu} \quad \text{for } 0 < \nu \leq 1.$$

In order to obtain approximations with this rate for $\nu > 1$ we may use extrapolated regularization [9]. That is, for a given $\beta > 0$ we set

$$U_{\beta}^{(0)}(t) = (t + \beta^{-1})^{-1}$$

and define Richardson extrapolants by

$$U_{\beta}^{(j)}(t) = (2^j U_{2\beta}^{(j-1)}(t) - U_{\beta}^{(j-1)}(t)) / (2^j - 1), \\ j = 1, 2, \dots$$

It is not difficult to show (see [9, lemma 2.1]) that for $k = 0, 1, 2, \dots$

$$t^{k+1} |1 - t U_{\beta}^{(k)}(t)| = \prod_{i=0}^k \left(\frac{t}{2^i \beta t + 1} \right) \\ \leq \beta^{-k-1}$$

Therefore, for the k th extrapolant we may apply Theorem 7 with $\omega(\beta, k) = \beta^{-k-1}$, $k = 1, 2, \dots$, to obtain the rate $\beta^{-k+1/2}$ (see [9, Theorem 3.2]).

We now consider some iterative regularization methods. Below, α will be a parameter satisfying $0 < \alpha < 2 \|T\|^{-2}$.

If the functions $U_n(t)$, $n = 0, 1, 2, \dots$ are defined by

$$U_n(t) = \alpha \sum_{k=0}^n (1 - \alpha t)^k$$

then (6) and (7) are satisfied and one can show that

$$n^{\nu} t^{\nu} |1 - t U_n(t)| = n^{\nu} t^{\nu} |1 - \alpha t|^{n+1}$$

is uniformly bounded. From this we find that the rate of convergence of the iterative process

$$x_0 = \alpha T^* b, \quad x_{n+1} = (I - \alpha \tilde{T}) x_n + \alpha T^* b$$

is determined by the function $\omega(n, \nu) = n^{-\nu}$.

Newton's method for approximating t^{-1} leads to the sequence of functions defined by

$$U_0(t) = \alpha, U_{n+1}(t) = U_n(t)(2 - tU_n(t)).$$

For this sequence of functions it is not difficult to see that

$$t^\nu |1 - tU_n(t)| = O(2^{-\nu n}) \quad \text{for } \nu > 0.$$

Therefore the rate of convergence of the corresponding iterative method is determined by the function $\omega(n, \nu) = 2^{-\nu n}$.

Showalter and Ben-Israel [20] have extrapolated on the previous method to obtain methods with a higher rate of convergence. For a positive integer $p \geq 2$ they define the hyperpower methods in terms of the sequence

$$U_0(t) = \alpha, U_{n+1}(t) = U_n(t) \sum_{k=0}^{p-1} (1 - tU_n(t))^k.$$

For these methods the results above may be used to obtain the convergence rate $O(p^{-\nu n})$.

In [11] Lardy considered the approximations

$$x_0 = 0, \quad \tilde{T}x_n + x_n = x_{n-1} + T^*b, \quad n = 1, 2, \dots$$

to T^*b , where T is an unbounded operator. We may apply the results above in the case of a bounded operator if we define the functions U_n by

$$U_n(t) = \sum_{k=1}^n (t+1)^{-k}.$$

One can verify, as in the first iterative example above, that the function $\omega(n, \nu) = n^{-\nu}$ determines a rate of convergence.

The iterative method

$$x_0 = T^*b, \quad x_{n+1} = x_n + (T^*b - \tilde{T}x_n)/(n+2),$$

was investigated in [10]. The appropriate functions U_n are given by

$$U_n(t) = \sum_{k=0}^n (k+1)^{-1} \prod_{j=0}^{k-1} (1 - t/(1+j)).$$

This leads to the iterative method

$$x_0 = T^*b, \quad x_{n+1} = x_n + (T^*b - \tilde{T}x_n)/(n+2).$$

Following the analysis given in [10] one can show that the rate of convergence of this method is governed by the function $\omega(n, \nu) = (\log n)^{-\nu}$.

REFERENCES

1. D.R. Audley and D.A. Lee, Considerations related to ill-posed and well-posed problems in system identification, ARL TR 73-D196, December, 1973.
2. E.C. Bullard and R.I.B. Cooper, The determination of masses necessary to produce a given gravitational field, Proc. Royal Soc. A 194(1968), 332-347.
3. J. Cullum, Numerical differentiation and regularization, SIAM J. Numer. Anal. 8(1971), 254-265.
4. T.S. Fong, On the problem of optimal antenna aperture distribution, J. Franklin Inst. 283(1967), 235-249.
5. J.N. Franklin, On Tikhonov's method for ill-posed problems, Math. Comp. 28(1974), 889-907.
6. R. Gorenflo (ed.), Inkorrekt Gestellte Probleme I, Preprint 58/77, Fachbereich Mathematik, Freie Universität Berlin, Berlin, 1977.
7. C.W. Groetsch, Generalized Inverses of Linear Operators: Representation and Approximation, Dekker, New York, 1977.
8. C.W. Groetsch, On rates of convergence for approximations to the generalized inverse, J. Numer. Func. Anal. Opt., to appear.
9. C.W. Groetsch and J.T. King, Extrapolation and the method of regularization for generalized inverses, J. Approx. Th., to appear.
10. C.W. Groetsch and B.J. Jacobs, Iterative methods for generalized inverses based on functional interpolation, in "Recent Applications of Generalized Inverses" (M.Z. Nashed, ed.), Pitman, London, to appear.
11. L.J. Lardy, A series representation for the generalized inverse of a closed linear operator, Tech. Rep. 74-18, Dept. of Math., Univ. of Maryland, April, 1974.
12. M.M. Lavrentiev, Some Improperly Posed Problems of Mathematical Physics, Springer-Verlag, Berlin, 1967.
13. D.A. Lee, On the determination of molecular weight distributions from sedimentation-diffusion equilibrium data at a single rotor speed, J. Polymer Sci. 8(1970), 1039-1056.

14. D.A. Lee et al., Some practical aspects of the treatment of ill-posed problems by regularization, ARL TR 75-0022, February, 1975.
15. M.Z. Nashed (ed.), Generalized Inverses and Applications, Academic Press, New York, 1976.
16. W.L. Perry, On the Bojarski-Lewis inverse scattering method, IEEE Trans. Antennas Prop. 6(1974), 826-829.
17. W.L. Perry, Approximate solution of inverse problems with piecewise continuous solutions, Radio Sci. 12(1977), 637-642.
18. S.W. Provencher, Numerical solution of linear integral equations of the first kind. Calculation of molecular weight distributions from sedimentation equilibrium data, J. Chem. Phys. 46(1967), 3229-3236.
19. D. Showalter, Representation and computation of the pseudoinverse, Proc. Amer. Math. Soc. 18(1967), 584-586.
20. D. Showalter and A. Ben-Israel, Representation and computation of the generalized inverse of a bounded linear operator between two Hilbert spaces, Accad. Naz. dei Lincei 48(1970), 184-194.
21. K. Singh and B. Paul, A method of solving ill-posed integral equations of the first kind, Comp. Meth. Appl. Mech. Eng. 2(1973), 339-348.
22. O.N. Strand and E.R. Westwater, Statistical estimation of the numerical solution of a Fredholm integral equation of the first kind, J.A.C.M. 15(1968), 100-114.
23. A.N. Tikhonov and V.Y. Arsenin, Solutions of Ill-Posed Problems, Wiley, New York, 1977.
24. V.F. Turchin et al., The use of mathematical-statistics methods in the solution of incorrectly posed problems, Sov. Phys. Uspekhi 13(1971), 681-703.

UNCONDITIONALLY STABLE SECOND ORDER ACCURATE METHOD
FOR TRANSONIC FLUTTER CALCULATIONS

Czeslaw P. Kentzer
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907

AUGUST 1978

UNCONDITIONALLY STABLE SECOND ORDER ACCURATE METHOD
FOR TRANSONIC FLUTTER CALCULATIONS

Czeslaw P. Kentzer*

Purdue University, West Lafayette, Indiana

Abstract

An implicit finite-difference method was developed for solving the full time-dependent potential equation of aerodynamics. The method was shown to be unconditionally stable and of second order accuracy in time and space. The intended use of the method is in a simultaneous solution of the aerodynamic and structural equations under transonic flutter conditions. The method was programmed and numerical results indicate the usual sensitivity of high order methods to disturbances produced by a shock wave. Several recommendations are made to remedy the situation. Among them is a suggestion on the correct use of the tri-diagonal algorithm of the Alternating Directions Implicit method in transonic flows where the domain of dependence of the solution changes as a function of the local Mach number. Further development of the method is needed to render it practical for flutter work.

* Assoc. Professor, School of Aeronautics & Astronautics, USAF-ASEE
Summer Faculty Research Fellow, Wright-Patterson Air Force Base, OH

LIST OF SYMBOLS

a	adiabatic speed of sound
A	difference operator, Equation (5)
A_j, B_j, C_j, D_j	coefficients of the tridiagonal matrix, Equation (10)
c	airfoil chord
C_n	Courant number
i	$\sqrt{-1}$
m	nondimensional wavelength, $\lambda/\Delta x$
M	Mach number
n	nondimensional wavelength, $\lambda/\Delta y$
p	pressure
Q, Q_x, Q_y	difference operators, Equation (5)
R, R_x, R_y	difference operators, Equation (5)
t	time
u	x-component of velocity
v	y-component of velocity
w	transformed amplification factor, $(\xi-1)/(\xi+1)$
x	streamwise Cartesian coordinate
x_j	auxiliary function, Equation (9)
y	normal Cartesian coordinate

LIST OF SYMBOLS (CONT'D)

α	$\Delta t \times$ complex frequency
β	$\Delta x \times$ x-component of wave number
γ	$\Delta y \times$ y-component of wave number
Γ_x	$\Delta t / \Delta x$
Γ_y	$\Delta t / \Delta y$
Δ	increment
ϵ	artificial dissipation function, Equation (6)
ζ	$e^{i\gamma}$
η	$e^{i\beta}$
λ	wavelength
ξ	$e^{i\alpha}$
ρ	mass density
σ	dissipation parameter, Equation (3)
ϕ	velocity potential

Subscripts

j	index denoting subdivisions along x-axis
k	index denoting subdivisions along y-axis
L	evaluated at lower surface
m	mean value
U	evaluated at upper surface
∞	free stream value

Superscript

n	index denoting subdivisions along t-axis
---	--

TABLE OF CONTENTS

<u>SECTION</u>		<u>PAGE</u>
I	INTRODUCTION	1
II	ALTERNATING DIRECTIONS IMPLICIT SCHEME	3
III	STABILITY ANALYSIS	7
IV	THE COMPUTATIONAL FORM OF THE FINITE DIFFERENCE SCHEME	13
V	THE KUTTA AND WAKE CONDITIONS	16
VI	RECOMMENDATIONS	20

FOREWORD

This report is a result of work carried out in the Optimization Group, Analysis and Optimization Branch, Structural Mechanics Division, Air Force Flight Dynamics Laboratory. It was performed by Dr. Czeslaw P. Kentzer, a Faculty Fellow at the Ohio State University in the USAF-ASEE Summer Faculty Program (WPAFB), June through August, 1978.

SECTION I

INTRODUCTION

The recent interest in transonic flutter created a demand for more accurate and efficient computational procedures for unsteady transonic aerodynamics. The present work is an attempt to develop such a procedure subject to the requirements that the numerical scheme be: (1) an approximation to the full potential equation, (2) of second order accuracy in both time and space, (3) applicable to high frequencies, (4) not limited to thin bodies, (5) easily extendable to three-dimensional wings and, (6) be unconditionally stable.

Earlier attempts at transonic flow computations, as reported in literature, lead to compromises so that not all of the above six conditions were satisfied. The difficulties of simultaneously meeting all these requirements have been partially overcome in the present work by application of central differencing in conjunction with the method of factorization (the Alternating Directions Implicit method or A.D.I.)

The computational method presented here requires further refinements and modifications. Computed results show poor shock capturing ability with flow field fluctuations characteristic of higher order methods. Without an artificially added numerical dis-

sipation the method is shown to be unconditionally neutrally (i.e. marginally) stable and of second order accuracy. Thus all six requirements have been met; however, the method does not appear at the present stage of development to be suitable for transonic calculations and several recommendations are made in SECTION VI to remedy the obvious shortcomings.

SECTION II gives a general procedure for deriving a factored scheme for the potential equation of aerodynamics which is hyperbolic and of second order in both time and space. In SECTION III the numerical stability of the linearized (constant coefficients) finite-difference equation is considered, while some computational aspects of the alternating directions method are discussed in SECTION IV. The treatment of the Kutta trailing edge and of the wake conditions is explained in SECTION V.

The main contributions of the work reported here are the development of a numerical scheme of high accuracy, both in time and in space, applicable to the full time-dependent potential equation not limited to thin bodies or low frequencies, and an extension of the tri-diagonal algorithm for integration in alternating directions to integration across the wake and across the imbedded supersonic zone. A novel idea of a type-dependent use of the method of Alternating Directions is presented in SECTION VI.

SECTION II

ALTERNATING DIRECTIONS IMPLICIT SCHEME

We shall consider the full potential equation for two-dimensional irrotational flows,

$$\begin{aligned} \frac{\partial^2 \phi}{\partial t^2} + 2(u \frac{\partial^2 \phi}{\partial x \partial t} + v \frac{\partial^2 \phi}{\partial y \partial t}) + u^2 \frac{\partial^2 \phi}{\partial x^2} + 2uv \frac{\partial^2 \phi}{\partial x \partial y} + v^2 \frac{\partial^2 \phi}{\partial y^2} \\ = a^2 (\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2}), \end{aligned} \quad (1)$$

nondimensionalized by referring u, v, a to U_∞ , x, y to chord c , the velocity potential ϕ to $U_\infty c$, and time t to c/U_∞ , where

$$a^2 = 1/M_\infty^2 - (\gamma - 1) [\frac{\partial \phi}{\partial t} + \frac{1}{2}(u^2 + v^2 - 1)]$$

is the nondimensional speed of sound, and $u = \partial \phi / \partial x$, $v = \partial \phi / \partial y$.

Introducing centered time differences, and with $t = n\Delta t$, $\phi(t) = \phi(n\Delta t) = \phi^n$, we have

$$\begin{aligned} \phi^{n+1} - 2\phi^n + \phi^{n-1} + \Delta t (u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y}) (\phi^{n+1} - \phi^{n-1}) \\ + \Delta t^2 (u^2 \frac{\partial^2}{\partial x^2} + 2uv \frac{\partial^2}{\partial x \partial y} + v^2 \frac{\partial^2}{\partial y^2}) \phi^n \\ = \Delta t^2 a^2 (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}) \phi^n + O(\Delta t^4). \end{aligned}$$

This may be rewritten as

$$\begin{aligned}
& \{1 + \Delta t(u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y})\} \phi^{n+1} \\
& - \{2 - 2\Delta t^2 uv \frac{\partial^2}{\partial x \partial y} + \Delta t^2 [(a^2 - u^2) \frac{\partial^2}{\partial x^2} + (a^2 - v^2) \frac{\partial^2}{\partial y^2}]\} \phi^n \\
& + \{1 - \Delta t(u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y})\} \phi^{n-1} = O(\Delta t^4).
\end{aligned}$$

Without affecting the order of accuracy, we add to the above the term (reasons for doing so are explained in SECTION IV)

$$- \frac{1}{4} \Delta t^2 [(a^2 + u^2) \frac{\partial^2}{\partial x^2} + (a^2 + v^2) \frac{\partial^2}{\partial y^2}] (\phi^{n+1} - 2\phi^n + \phi^{n-1}) = O(\Delta t^4)$$

and obtain

$$\begin{aligned}
& \{1 + \Delta t(u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y}) - \frac{\Delta t^2}{4} [(a^2 + u^2) \frac{\partial^2}{\partial x^2} + (a^2 + v^2) \frac{\partial^2}{\partial y^2}]\} \phi^{n+1} \\
& - 2\{1 - \Delta t^2 uv \frac{\partial^2}{\partial x \partial y} + \frac{\Delta t^2}{4} [(a^2 - 3u^2) \frac{\partial^2}{\partial x^2} + (a^2 - 3v^2) \frac{\partial^2}{\partial y^2}]\} \phi^n \\
& + \{1 - \Delta t(u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y}) - \frac{\Delta t^2}{4} [(a^2 + u^2) \frac{\partial^2}{\partial x^2} + (a^2 + v^2) \frac{\partial^2}{\partial y^2}]\} \phi^{n-1} \\
& = O(\Delta t^4). \tag{2}
\end{aligned}$$

Now consider the following expressions, the first of which provides some damping (dissipation) and the second completes the products of differential operators:

$$\begin{aligned}
& - 2\sigma \Delta t^2 \Delta x^2 \Delta y^2 (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}) \frac{\partial \phi}{\partial t} \\
& = - 2\sigma \Delta t \Delta x^2 \Delta y^2 (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}) (\phi^{n+1} - \phi^{n-1}) + O(\Delta t^4), \tag{3}
\end{aligned}$$

and

$$\begin{aligned}
& \Delta t^2 uv \frac{\partial^2}{\partial x \partial y} (\phi^{n+1} - 2\phi^n + \phi^{n-1}) + \frac{\Delta t^4}{16} (a^2 + u^2) (a^2 + v^2) \frac{\partial^4}{\partial x^2 \partial y^2} (\phi^{n+1} + \phi^{n-1}) \\
& - \Delta t^2 \sigma \Delta x^2 \Delta y^2 \left(u \frac{\partial^3}{\partial x \partial y^2} + v \frac{\partial^3}{\partial x^2 \partial y} + \sigma \Delta x^2 \Delta y^2 \frac{\partial^4}{\partial x^2 \partial y^2} \right) (\phi^{n+1} + \phi^{n-1}) \\
& - \frac{\Delta t^3}{4} \left\{ u (a^2 + v^2) \frac{\partial^3}{\partial x \partial y^2} + v (a^2 + u^2) \frac{\partial^3}{\partial x^2 \partial y} \right\} (\phi^{n+1} - \phi^{n-1}). \quad (4)
\end{aligned}$$

If $\sigma \Delta t \Delta x^2 \Delta y^2 = O(\Delta t^4)$, then both expressions (3) and (4) are $O(\Delta t^4)$.

Adding (3) and (4) to Equ. (2) permits one to factor the spatial operators as follows:

$$\begin{aligned}
& (Q_x - \sigma \Delta t \Delta x^2 \Delta y^2 \frac{\partial^2}{\partial x^2}) (Q_y - \sigma \Delta t \Delta x^2 \Delta y^2 \frac{\partial^2}{\partial y^2}) \phi^{n+1} - A \phi^n \\
& + (R_x + \sigma \Delta t \Delta x^2 \Delta y^2 \frac{\partial^2}{\partial x^2}) (R_y + \sigma \Delta t \Delta x^2 \Delta y^2 \frac{\partial^2}{\partial y^2}) \phi^{n-1} = 0 \quad (5)
\end{aligned}$$

where

$$\begin{aligned}
Q_x &= 1 - \frac{\Delta t^2}{4} (a^2 + u^2) \frac{\partial^2}{\partial x^2} + u \Delta t \frac{\partial}{\partial x}, \\
R_x &= 1 - \frac{\Delta t^2}{4} (a^2 + u^2) \frac{\partial^2}{\partial x^2} - u \Delta t \frac{\partial}{\partial x}, \\
Q_y &= 1 - \frac{\Delta t^2}{4} (a^2 + v^2) \frac{\partial^2}{\partial y^2} + v \Delta t \frac{\partial}{\partial y}, \\
R_y &= 1 - \frac{\Delta t^2}{4} (a^2 + v^2) \frac{\partial^2}{\partial y^2} - v \Delta t \frac{\partial}{\partial y}, \\
A &= 2 + \frac{\Delta t^2}{2} [(a^2 - 3u^2 \delta_x) \frac{\partial^2}{\partial x^2} + (a^2 - 3v^2 \delta_y) \frac{\partial^2}{\partial y^2}],
\end{aligned}$$

and where the symbols δ_x and δ_y indicate a further possible modification resulting in increased numerical dissipation. We may, e.g., use here finite difference averaging operators which replace

the value of the mesh function at a given point by a weighted sum (average) of its neighbors.

In order to maintain a second order temporal accuracy, the dissipation parameter σ should be kept proportional to $\Delta t^2 / \Delta x^2 \Delta y^2$. Spatial accuracy depends on the approximation to the spatial partial derivatives. With centered spatial differences second order accuracy with respect to space is assured.

Implementation of the Alternating Directions method of solution of the factored scheme (5) is treated in SECTION IV.

SECTION III

STABILITY ANALYSIS

We shall limit the study of stability to the case of linear difference equations, that is, to the finite-difference analog of Equ. (5) with u , v , and a kept constant.

A linear difference equation admits a fundamental solution in the form

$$\phi_{jk}^n = \phi(n\Delta t, j\Delta x, k\Delta y) = \phi_0 e^{i(n\alpha + j\beta + k\gamma)}$$

where $\alpha/\Delta t$ = complex frequency, and where $\beta/\Delta x$ and $\gamma/\Delta y$ are the components of the wave number vector. We note the following shift property of the fundamental solutions,

$$\phi_{j+J, k+K}^{n+N} = (e^{i\alpha})^N (e^{i\beta})^J (e^{i\gamma})^K \phi_{jk}^n = \xi^N \eta^J \zeta^K \phi_{jk}^n.$$

For stability in the von Neumann sense we require that $|\xi| \leq 1$. Further, we observe that the amplification factor ξ and the unit vectors η and ζ are eigenvalues of the shift operators, and that finite differences are formed by taking differences of the mesh function $\phi(n\Delta t, j\Delta x, k\Delta y)$ with its arguments shifted (increased or decreased by integer multiples of the increments Δt , Δx , Δy). Thus $\phi_{jk}^{n\pm 1} = \xi^{\pm 1} \phi_{jk}^n$, and the centered spatial differences are

$$\frac{\partial \phi}{\partial x} \approx \frac{1}{2\Delta x} (\phi_{j+1, k}^n - \phi_{j-1, k}^n) = \frac{\eta - \eta^{-1}}{2\Delta x} \phi_{jk}^n = \left(\frac{i}{\Delta x} \sin \beta \right) \phi_{jk}^n,$$

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{1}{\Delta x^2} (\phi_{j+1,k}^n - 2\phi_{jk}^n + \phi_{j-1,k}^n) = \frac{n-2+n}{\Delta x^2} \phi_{jk}^n = -\frac{2}{\Delta x^2} (1-\cos\beta) \phi_{jk}^n,$$

and similarly,

$$\frac{\partial^2 \phi}{\partial y^2} = -\frac{2}{\Delta y^2} (1-\cos\gamma) \phi_{jk}^n, \quad \frac{\partial \phi}{\partial y} = \left(\frac{i}{\Delta y} \sin\gamma\right) \phi_{jk}^n,$$

$$\delta_x = \frac{1}{4} (\phi_{j+1,k}^n + 2\phi_{jk}^n + \phi_{j-1,k}^n) = \frac{1}{2} (1+\cos\beta) \phi_{jk}^n,$$

$$\delta_y = \frac{1}{4} (\phi_{j,k+1}^n + 2\phi_{jk}^n + \phi_{j,k-1}^n) = \frac{1}{2} (1+\cos\gamma) \phi_{jk}^n.$$

Substituting the fundamental solution into Equ. (5) we have with the notation $\Gamma_x \equiv \Delta t/\Delta x$, $\Gamma_y \equiv \Delta t/\Delta y$, and with

$$Q_x = 1 + \frac{1}{2} \Gamma_x^2 (a^2 + u^2) (1-\cos\beta) + iu\Gamma_x \sin\beta,$$

$$Q_y = 1 + \frac{1}{2} \Gamma_y^2 (a^2 + v^2) (1-\cos\gamma) + iv\Gamma_y \sin\gamma,$$

$$R_x = 1 + \frac{1}{2} \Gamma_x^2 (a^2 + u^2) (1-\cos\beta) - iu\Gamma_x \sin\beta = \bar{Q}_x,$$

$$R_y = 1 + \frac{1}{2} \Gamma_y^2 (a^2 + v^2) (1-\cos\gamma) - iv\Gamma_y \sin\gamma = \bar{Q}_y,$$

$$A = 2\left\{1 - \frac{1}{2} \Gamma_x^2 \left[a^2 - \frac{3}{2} u^2 (1+\cos\beta)\right] (1-\cos\beta) - \frac{1}{2} \Gamma_y^2 \left[a^2 - \frac{3}{2} v^2 (1+\cos\gamma)\right] (1-\cos\gamma)\right\},$$

$$(Q + \varepsilon)\xi - A + (R - \bar{\varepsilon})\xi^{-1} = (Q + \varepsilon)\xi - A + (\bar{Q} - \bar{\varepsilon})\xi^{-1} = 0. \quad (6)$$

where the overbar indicates a complex conjugate, and where

$$Q = Q_x Q_y + 4\sigma^2 \Delta t^2 \Delta x^2 \Delta y^2 (1-\cos\beta) (1-\cos\gamma) = Q_r + i Q_i,$$

$$\varepsilon = 2\sigma \Delta t [Q_x \Delta y^2 (1-\cos\gamma) + Q_y \Delta x^2 (1-\cos\beta)] = \varepsilon_r + i\varepsilon_i.$$

Equation (6) is a polynomial with complex coefficients and of second degree in ξ . The coefficients, and therefore also the roots of the polynomial, depend on eight parameters, viz. on Δt , Δx , Δy , u , v , a^2 and the angles β and γ . We are unable to solve the complex polynomial for ξ in a closed form. However, we may make use of the standard techniques of the stability theory. We first multiply the quadratic (6) by its complex conjugate polynomial,

$$(\bar{Q} + \bar{\epsilon})\xi^2 - A\xi + (Q - \epsilon) = 0,$$

to obtain a polynomial with real coefficients,

$$[Q\bar{Q} + \bar{Q}\epsilon + Q\bar{\epsilon} + \epsilon\bar{\epsilon}]\xi^4 - [Q + \bar{Q} + \epsilon + \bar{\epsilon}]A\xi^3 + [A^2 + Q^2 + \bar{Q}^2 - (\epsilon^2 + \bar{\epsilon}^2)]\xi^2 - [Q + \bar{Q} - (\epsilon + \bar{\epsilon})]A\xi + [Q\bar{Q} - Q\bar{\epsilon} - \bar{Q}\epsilon + \epsilon\bar{\epsilon}] = 0,$$

and then transform the ξ -plane, using $\xi = (1+w)/(1-w)$, into the w -plane where the stability condition, $|\xi| \leq 1$, transforms to $\text{Re}(w) \leq 0$. The transformed polynomial is of the form

$$a_0 w^4 + a_1 w^3 + a_2 w^2 + a_3 w + a_4 = 0 \quad (7)$$

where $a_0 = (A + Q + \bar{Q})^2 + 4\epsilon\bar{\epsilon} > 0$,

$$a_1 = 8[\bar{Q}\epsilon + Q\bar{\epsilon} + \frac{1}{2}A(\epsilon + \bar{\epsilon})],$$

$$a_2 = 2[A^2 + (Q + \bar{Q})^2 + 4Q\bar{Q} + 4\epsilon\bar{\epsilon} + 4\epsilon\bar{\epsilon}] > 0,$$

$$a_3 = 8[\bar{Q}\epsilon + Q\bar{\epsilon} - \frac{1}{2}A(\epsilon + \bar{\epsilon})],$$

$$a_4 = [A - (Q + \bar{Q})]^2 + 4\epsilon\bar{\epsilon} > 0.$$

The numerical dissipation term, ϵ , which was added to Equ. (2), has the effect of modifying the unconditional neutral (marginal)

stability, for, with $\varepsilon = 0$ the Routh-Hurwitz sufficient conditions for stability are in the present case:

$$\begin{aligned}\Delta_0 &= a_0 > 0, \quad \Delta_1 = a_1 = 0, \quad \Delta_2 = a_1 a_2 - a_0 a_3 = 0, \\ \Delta_3 &= a_1 a_2 a_3 - a_0 (a_1 a_4 - a_3^2) = 0, \quad \Delta_4 = a_4 \Delta_3 = 0.\end{aligned}$$

Thus $\text{Re}(w) = 0$ or $|\xi| = 1$ unconditionally with zero dissipation ($\varepsilon = 0$) and the scheme has no damping or amplification.

The addition of the damping term, while it lowers the order of temporal accuracy (unless $\sigma \propto \Delta t^2 / \Delta x^2 \Delta y^2$), modifies the coefficients a_0, \dots, a_4 of the transformed polynomial by an addition of quadratic terms $\varepsilon_r^2, \varepsilon_i^2$ to the even-numbered coefficients while making the odd-numbered coefficients proportional to the magnitude of ε . Reversing the sign of ε changes the sign of a_1 and a_3 . Since the necessary condition for $\text{Re}(w) < 0$ is that all the coefficients, a_0, \dots, a_4 , be positive, the sign of ε (e.g. that of σ in Equ. (7)) must be chosen so that $a_1 + a_3 = 16(Q_r \varepsilon_r + Q_i \varepsilon_i)$ be positive; but this does not yet guarantee that both a_1 and a_3 are separately greater than zero.

Due to the algebraic complexity the full set of inequalities, the necessary stability conditions $a_0 > 0, a_1 > 0, \dots, a_4 > 0$ and the Routh-Hurwitz sufficient conditions $\Delta_0 > 0, \Delta_1 > 0, \dots, \Delta_4 > 0$ could not be given in a closed form. We note, however, that the even numbered coefficients a_0, a_2 , and a_4 in Equ. (7) are positive

definite. Consequently the two inequalities $a_1 > 0$ and $a_3 > 0$ are of primary concern to us.

Upon a closer examination of the expression for a_3 the following conclusions were reached. With $\beta = 2\pi\Delta x/\lambda$, $\lambda = m\Delta x$ = wavelength of disturbance, $m = 2, 3, 4, \dots$, thus $\beta = 2\pi/m$; similarly, $\gamma = 2\pi/n$ where $n = 2, 3, 4, \dots$. Short waves (e.g., $m=2$ and/or $n=2$) are definitely always stable and highly damped, the latter property being very desirable as it prevents the appearance of "wiggles" and the "up-and-down" oscillations of the numerical solution in presence of shock waves. Long waves, $m, n \rightarrow \infty$, are only marginally stable. Again, this is desirable. However, certain combinations of large but finite wavelengths may change the sign of a_3 depending on Δt , the Mach number and the mesh aspect ratio $Ar = \Delta y/\Delta x$. Numerical evaluation of the roots of the complex polynomial (6) confirmed this conclusion. Instability, $|\xi| > 1$, occurs only at supersonic points and for long waves, and it seems to require the combination of medium and long wavelengths, e.g., $8\Delta x$ and $256\Delta y$, or $16\Delta x$ and $128\Delta y$, etc. These conclusions were confirmed by supercritical flow calculations on a 56×32 point mesh where the shortest observed oscillations had a wavelength of $8\Delta x$. These oscillations disappeared on a coarser mesh due to the boundary conditions. This is a classical case of a "long wave instability" which may occur as the mesh is continuously refined until the unstable long waves

fit between the boundaries, that is, when the number of subdivisions of the computational mesh equals the half-wavelength divided by the mesh spacing, $\frac{1}{2}m$ or $\frac{1}{2}n$ in our notation.

Due to the marginal damping and/or instability of the long waves, the shock wave thickness is large, on the order of the half wavelength of the waves with largest amplitude (that is, of those having the least damping). The dispersion properties of such waves should be studied with the objective of modifying the computational scheme so as to improve the situation.

The obvious changes in the computational scheme that offer hope of improving the performance of the method are the conservative differencing and the type-dependent differencing. Since the shock wave creates the major disturbance in the solution, discrete handling of the shock waves would remove most of the difficulties encountered in the present work.

SECTION IV

THE COMPUTATIONAL FORM OF THE FINITE DIFFERENCE SCHEME

Using centered spatial differences we put the scheme (5) into a form emphasizing one-dimensional character of the Alternating Directions Implicit method (ADI). Let

$$G_{jk} = \phi_{jk}^{n-1} - \frac{\Delta t}{\Delta y} \{ \frac{1}{2} v (\phi_{j,k+1}^{n-1} - \phi_{j,k-1}^{n-1}) - [\sigma \Delta x^2 \Delta y - \frac{1}{4} \frac{\Delta t}{\Delta y} (a^2 + v^2)] (\phi_{j,k+1}^{n-1} - 2\phi_{j,k}^{n-1} + \phi_{j,k-1}^{n-1}) \},$$

$$F_{jk} = G_{jk} - \frac{\Delta t}{\Delta x} \{ \frac{1}{2} u (G_{j+1,k} - G_{j-1,k}) - [\sigma \Delta x \Delta y^2 - \frac{1}{4} \frac{\Delta t}{\Delta x} (a^2 + u^2)] (G_{j+1,k} - 2G_{j,k} + G_{j-1,k}) \},$$

$$H_{jk} = 2\phi_{jk}^n + \frac{a^2 \Delta t^2}{2\Delta x^2} (\phi_{j+1,k}^n - 2\phi_{j,k}^n + \phi_{j-1,k}^n) - \frac{3u^2 \Delta t^2}{8\Delta x^2} (\phi_{j+2,k}^n - 2\phi_{j,k}^n + \phi_{j-2,k}^n) + \frac{a^2 \Delta t^2}{2\Delta y^2} (\phi_{j,k+1}^n - 2\phi_{j,k}^n + \phi_{j,k-1}^n) - \frac{3v^2 \Delta t^2}{8\Delta y^2} (\phi_{j,k+2}^n - 2\phi_{j,k}^n + \phi_{j,k-2}^n).$$

With the solution given at the two time levels, $t = (n-1)\Delta t$ and $t = n\Delta t$, the functions G_{jk} , F_{jk} , and H_{jk} are known. Equation (5) takes then the form

$$x_{jk} + \frac{\Delta t}{\Delta x} \{ \frac{1}{2} u (x_{j+1,k} - x_{j-1,k}) - [\sigma \Delta x \Delta y^2 - \frac{1}{4} \frac{\Delta t}{\Delta x} (a^2 + u^2)] (x_{j+1,k} - 2x_{j,k} + x_{j-1,k}) \} = H_{jk} - F_{jk} \quad (8)$$

where the auxiliary function X_{jk} is defined by

$$X_{jk} = \phi_{j,k}^{n+1} + \frac{\Delta t}{\Delta y} \left\{ \frac{1}{2} v (\phi_{j,k+1}^{n+1} - \phi_{j,k-1}^{n+1}) - [\sigma \Delta x^2 \Delta y + \frac{\Delta t}{4} (a^2 + v^2)] (\phi_{j,k+1}^{n+1} - 2\phi_{j,k}^{n+1} + \phi_{j,k-1}^{n+1}) \right\}. \quad (9)$$

Equation (8) contains the unknown X evaluated only at three points allowing for the use of the efficient tri-diagonal matrix inversion algorithm. We write (8) as

$$-A_j X_{j+1,k} + B_j X_{j,k} - C_j X_{j-1,k} = D_j \quad (10)$$

where $A_j = \sigma \Delta t \Delta y^2 + \frac{\Delta t^2}{4 \Delta x^2} (a^2 + u^2) - \frac{u \Delta t}{2 \Delta x},$

$$B_j = 1 + 2\sigma \Delta t \Delta y^2 + \frac{\Delta t^2}{2 \Delta x^2} (a^2 + u^2),$$

$$C_j = \sigma \Delta t \Delta y^2 + \frac{\Delta t^2}{4 \Delta x^2} (a^2 + u^2) + \frac{u \Delta t}{2 \Delta x},$$

$$D_j = H_{jk} - F_{jk}.$$

The solution of Equ. (10) for X_{jk} is to be followed immediately by the application of the tri-diagonal algorithm to Equ. (9) written as

$$-A_k \phi_{j,k+1}^{n+1} + B_k \phi_{j,k}^{n+1} - C_k \phi_{j,k-1}^{n+1} = D_k,$$

where $A_k = \sigma \Delta t \Delta x^2 + \frac{\Delta t^2}{4 \Delta y^2} (a^2 + v^2) - \frac{v \Delta t}{2 \Delta y},$

$$B_k = 1 + 2\sigma \Delta t \Delta x^2 + \frac{\Delta t^2}{2 \Delta y^2} (a^2 + v^2),$$

$$C_k = \sigma \Delta t \Delta x^2 + \frac{\Delta t^2}{4 \Delta y^2} (a^2 + v^2) + \frac{v \Delta t}{2 \Delta y}, \text{ and } D_k = X_{jk}.$$

We note at this point the advantage of the addition of the proper terms used to obtain Equ.(2) in SECTION II. If this were not done, the terms containing the factors (a^2+u^2) and (a^2+v^2) would be absent from the expressions for the coefficients A, B, C given above. The importance of retaining such terms follows from the properties of the tri-diagonal algorithm. In order that the round-off errors of the tri-diagonal solver not grow exponentially, one must require that $A > 0$, $B > 0$, $C > 0$ and $B > A+C$. Without the aforementioned terms the first inequality (or the third, depending on the direction of flow) could be violated especially for small dissipation parameter σ . Computations with $\sigma = 0$ would not be possible. With the addition of the proper terms leading to Equ.(2) the round-off errors will be bounded if we require that the Courant number C_n satisfy the inequality

$$C_n = a\Delta t \left\{ \frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} \right\}^{\frac{1}{2}} > \frac{2M}{1+M^2} .$$

This condition is sufficient and not unduly restrictive since $2M/(1+M^2) \leq 1$ and the objective of implicit method is to avoid the restriction of the Courant-Friedrichs-Levi stability criterion $C_n \leq 1$.

SECTION V

THE KUTTA AND WAKE CONDITIONS

The imposition of the Kutta trailing edge condition and the wake condition presents a special problem in the Alternating Direction Method, namely, that two conditions must be met simultaneously in the middle of the interval of integration in the direction normal to the wake.

In practice, the trailing edge is usually located between two mesh points so that the Kutta condition and the wake condition lead to the same physical requirement that, for all points downstream of the trailing edge, there be no pressure difference across the wake and that the mass flow across the wake be continuous. For simplicity, we take the wake as lying in the plane of the wing. Mathematically, the Kutta and wake conditions become

$$P_U = P_L, \quad (\partial\phi/\partial y)_U = (\partial\phi/\partial y)_L,$$

where the subscripts $()_U$ and $()_L$ denote sides of the wake corresponding to the upper and lower surfaces of the airfoil, respectively.

Using the exact relations for an irrotational flow,

$$\gamma p/\rho = \gamma p \frac{\gamma-1}{\gamma} = a^2 = a_\infty^2 - (\gamma-1) \left\{ \frac{\partial\phi}{\partial t} + \frac{1}{2} \left[\left(\frac{\partial\phi}{\partial x} \right)^2 + \left(\frac{\partial\phi}{\partial y} \right)^2 - 1 \right] \right\}$$

and the continuity of the normal component of velocity we obtain as a condition of continuity of pressure across the wake

$$\frac{\partial}{\partial t}(\phi_U - \phi_L) + \frac{1}{2}\left\{\left(\frac{\partial\phi}{\partial x}\right)_U^2 - \left(\frac{\partial\phi}{\partial x}\right)_L^2\right\} = 0.$$

This may be rewritten as

$$\left\{\frac{\partial}{\partial t} + \frac{1}{2}\left[\left(\frac{\partial\phi}{\partial x}\right)_U + \left(\frac{\partial\phi}{\partial x}\right)_L\right]\left(\frac{\partial}{\partial x}\right)\right\}(\phi_U - \phi_L) = 0. \quad (11)$$

Thus the jump in the potential, $\Delta\phi = \phi_U - \phi_L$, propagates along the wake with the instantaneous local mean velocity U_m ,

$$U_m = \frac{1}{2}\left\{\left(\frac{\partial\phi}{\partial x}\right)_U + \left(\frac{\partial\phi}{\partial x}\right)_L\right\},$$

and represents the vorticity shed from the trailing edge. The potential jump is, in general, a function of time and position in the case of oscillating airfoil, is constant in space in the low frequency approximation and constant in time in the steady flow case. The jump vanishes only for a nonlifting airfoil.

The continuity of the normal velocity component may be imposed conveniently by requiring that the adjusted potential, $\phi' = \phi$ for $y < 0$, $\phi' = \phi - \Delta\phi$ for $y > 0$, be continuous together with its first derivatives across the wake. This will facilitate handling of the boundary conditions in the y-direction step of the ADI method.

If the accuracy of the finite difference approximation to Equ.(11) is not of primary concern, one may finite-difference

Equ.(11) as follows,

$$\Delta\phi_{j,k}^{n+1} = \Delta\phi_{j,k}^n + \Delta t U_m (\Delta\phi_{j,k}^n - \Delta\phi_{j-1,k}^n) + O(\Delta t^2) + O(\Delta x^2). \quad (12)$$

Equation (12) provides accuracy of only first order in time and space. An improvement in accuracy may be realized using

- (a) an implicit scheme centered in space which may suffer from exponentially growing round-off errors when solved by the tri-diagonal inversion algorithm,

$$\begin{aligned} -\frac{1}{2}U_m \Delta t \Delta\phi_{j+1,k}^{n+1} + \Delta\phi_{j,k}^{n+1} + \frac{1}{2}U_m \Delta t \Delta\phi_{j-1,k}^{n+1} \\ = \Delta\phi_{j,k}^n + \frac{1}{2}U_m \Delta t (\Delta\phi_{j+1,k}^n - \Delta\phi_{j-1,k}^n) + O(\Delta t^4) + O(\Delta x^4), \end{aligned}$$

- (b) the explicit leap-frog scheme requiring three levels of storage,

$$\Delta\phi_{j,k}^{n+1} = \Delta\phi_{j,k}^{n-1} + U_m \Delta t (\Delta\phi_{j+1,k}^n - \Delta\phi_{j-1,k}^n) + O(\Delta t^4) + O(\Delta x^4).$$

Other widely used schemes could be employed here. In three dimensional flows similar technique could be used with the jump in the potential being carried unchanged along the vortex sheet with the mean velocity having now two components in the plane of the wing.

The treatment of the wake suggested here is only approximate since the wake does not remain in the plane of the wing. A discrete handling of the wake as a moving discontinuity, while it offers elegance and accuracy, would soon lead to complications

due to a natural instability of a vortex sheet which has a tendency to roll up into large vortex structures. Out of two evils we suggest to use the smaller.

SECTION VI

RECOMMENDATIONS

There are two obvious shortcomings of the proposed numerical scheme at this stage of its development, namely, the neutral stability with no artificial dissipation ($\sigma = 0$), and the poor resolution of shock waves. Further investigation of the effects of adding higher order dissipative terms without long-wave destabilization of the scheme should be carried out systematically with the objective of retaining a second order accuracy in both time and space while assuring stability independently of Mach number, mesh size and the coordinate system.

Conservative differencing is necessary to capture the shock wave motion properly. The present method should, therefore, be applied to the divergence (conservative) form of the potential equation,

$$\frac{\partial}{\partial t} \left[\frac{\partial \phi}{\partial t} + \frac{1}{2} (\nabla \phi)^2 \right] + \nabla \cdot \left\{ \left[\frac{\partial \phi}{\partial t} + \frac{1}{2} (\nabla \phi)^2 - \frac{a_\infty^2}{\gamma - 1} \right] \nabla \phi \right\} = 0,$$

which is equivalent to Equ. (1) when expanded. However, since finite difference operators applied to nonlinear terms do not commute, additional terms would appear in the finite difference equations (8) and (9). Conservative differencing conserves fluxes of conserved quantities and this reduces truncation errors across discontinuities which errors govern the amplitudes of the

oscillations of the solution at the shock.

Conservative differencing alone would not, however, improve handling of the shock waves sufficiently. A type-dependent scheme (upstream differencing at supersonic points) could be used to improve shock capturing and the shock wave resolution. Conservative differencing would not, per se, remedy the stability problem since the stability analysis of the linearized finite-difference equation would not be affected by conservative differencing.

The author is unaware of any numerical experiments and/or theoretical arguments for or against the following idea. In order to eliminate the unwanted downstream dependence in supersonic regions, the x-direction (streamwise) step of the ADI method could be broken up into three steps (the y-component of velocity will ordinarily remain subsonic). In a transonic flow the line $y = \text{constant}$ will cut the imbedded supersonic zone at two points. Customarily, this fact is ignored and two boundary conditions are used, one at the inlet ($x \rightarrow -\infty$) and one at the outlet ($x \rightarrow +\infty$) boundaries of the computational region. The tri-diagonal algorithm calculates the intermediate solution in the x-direction ADI step subject to these boundary conditions treating the problem as a typical two-point boundary value case. If, instead, the $y = \text{constant}$ line is divided into three segments, a subsonic region between $x = -\infty$ and the sonic line, followed by a supersonic region between the sonic line and the shock wave,

and another subsonic region between the shock and $x = +\infty$, then each of the subsonic regions represents a two-point boundary value problem solvable by the tri-diagonal algorithm, while in the supersonic region we need two initial conditions at the sonic line (e.g., the value of the potential and its x -derivative), and a recursion formula, essentially given by Equ. (10), would provide a solution not depending on the upstream boundary. The solution in the supersonic region would yield an initial condition for the continuation of the solution in the subsonic region downstream of the shock. A simple modification of the ADI procedure may thus accomplish what the upstream differencing technique had as a sole monopoly - namely, that of imposing correctly the rule of forbidden signals. Quite possibly, the type-dependent ADI algorithm proposed here could be used successfully in conjunction with central differencing thus preserving the second order spatial accuracy.

The above proposed type-dependent ADI method could be easily implemented in any existing program using the tri-diagonal algorithm. It modifies the standard use of the ADI method from its customary applications to elliptic problems to the solution of mixed and hyperbolic problems.

LARGE AMPLITUDE RESPONSE OF COMPLEX STRUCTURES
DUE TO HIGH INTENSITY NOISE

Chuh Mei
Department of Engineering Mechanics
University of Missouri - Rolla
Rolla, Missouri 65401

August 1978

AIR FORCE FLIGHT DYNAMICS LABORATORY
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

PREFACE

This report summarizes the results of the studies made on the nonlinear response of complex structural panels subjected to broadband random acoustic excitation by the author during his ten week stay at AFFDL/FBED, W-PAFB, OH 45433. The work was supported by the Air Force Office of Scientific Research through the USAF-ASEE Summer Faculty Research Program (W-PAFB), Contract F44620-76-C-0052, The Ohio State University, Columbus, OH 43210.

The author wishes to extend his great appreciation to Dr. Cecil D. Bailey, Director, USAF-ASEE Summer Faculty Program (W-PAFB) for giving him this unique opportunity of directly associating himself with the Air Force peers and thus becoming acquainted with the practical problems of mutual interest.

The author is extremely thankful to Mr. Howard L. Wolfe, Air Force Colleague, and Mr. Robert M. Bader, Chief, Structural Integrity Branch, Structural Mechanics Division, AFFDL, for providing him with every facility and help needed to carry out these investigations at the AFFDL and making his stay at W-PAFB as comfortable and fruitful as possible.

Special acknowledgement is due Mr. Ralph M. Shimovetz, Mr. Kenneth R. Wentz, Dr. V.B. Venkayya, and Mr. N.D. Wolf of AFFDL for many technical discussions. The last but not the least person the author would like to thank is Barbara Nickerson for her help in getting this manuscript typed.

4 August 1978

Chuh Mei

LARGE AMPLITUDE RESPONSE OF COMPLEX STRUCTURES DUE TO HIGH INTENSITY NOISE

Chuh Mei*
Department of Engineering Mechanics
University of Missouri - Rolla
Rolla, Missouri 65401

ABSTRACT

A constant problem of interest to the Air Force is the design of acoustically sound aircraft structural components. This is because sonic fatigue failures have resulted in unacceptable maintenance and inspection burdens associated with the operation of the aircraft. In some instances, sonic fatigue failures have resulted in major redesign efforts of structural components. Currently, the sonic fatigue design methods in practice are completely based on the linear or small deflection theory (Sonic Fatigue Design Guide for Military Aircraft, AFFDL-TR-74-112, for example). But, on the contrary, the test structural panels respond nonlinearly with large deflections at high intensity acoustic pressure levels. This large amplitude geometrical nonlinearity is identified to be the major factor that causes the enormous disagreement between the computed and the measured random responses. To improve the sonic fatigue design method, therefore, large deflection or nonlinear structure theory has to be employed in the analysis. This paper first gives a thorough survey-type review of various existing analytical and numerical methods on random excitations of nonlinear multidegree-of freedom systems, and an evaluation of these methods based on some realistic considerations from the point of view of their application to complex panel configurations of aircraft structure. These are the Fokker-Planck equation method, equivalent linearization technique, perturbation approach, finite difference method, finite element-equivalent linearization approach, etc. Then, a mathematical formulation, which is based on finite element method and equivalent linearization technique, for complex structural panels to high noise environment is developed. Statistical responses of nodal deflections and element strains using a single-mode approximation are given in terms of the linear frequency, spectral density of excitation pressure, equivalent linear frequency, and strain-deflection transformation matrices. Extension of the quasi-linearization method, which has been used successfully in analysis of large amplitude vibrations of complex structures, to the present nonlinear random excitation problem is also discussed briefly.

*Associate Professor.

CONTENTS

Section		Page
I	INTRODUCTION	1
II	EVIDENCE OF LARGE AMPLITUDE NONLINEARITY	4
III	REVIEW OF EXISTING APPROACHES ON NONLINEAR RANDOM VIBRATION	11
	1. Analytical Methods	11
	a. Fokker-Planck Equation Approach	
	b. Equivalent Linearization Approach	
	c. Perturbation Approach	
	d. Other Approximate Methods	
	2. Numerical Methods	14
	a. Finite Difference Approach	
	b. Finite Element Method	
IV	MATHEMATICAL FORMULATION	16
	1. Finite Element Representation	16
	2. Damping Representation	17
	3. Equivalent Linearization Approach	20
	4. Single-Mode Approximation and Mean-Square Responses	21
	5. Iteration Procedure and Flow Chart	25
	6. Estimation of Improvement	26
V	GEOMETRICAL STIFFNESS MATRICES	30
VI	QUASI-LINEARIZATION METHOD	33
VII	CONCLUSIONS	34
	1. Summary of Results	34
	2. Recommendations	34
	REFERENCES	36-39

NOMENCLATURE

$[c]$	Element damping matrix
$[C], [C]_s$	Damping matrices
$[C]$	Diagonal generalized damping matrix
e	Error of linearization or equation deficiency
$E[\]$	Operator denotes the mathematical expectation
$\{f\}$	Element nodal force vector
$\{F\}, \{F\}_s$	Nodal force vectors
g	Structural damping coefficient
$H(\Omega)$	Frequency response function
$[k]$	Element stiffness matrix
$[K], [K]_s$	Stiffness matrices
$[K]$	Diagonal generalized stiffness matrices
$[k^g]$	Element geometrical stiffness matrix
$[K^g], [K^g]_s$	Geometrical stiffness matrices
k^{eq}	Equivalent linear stiffness constant
$[m]$	Element consistent mass matrix
$[M], [M]_s$	Consistent mass matrices
$[M]$	Diagonal generalized mass matrix
$\{P\}$	Force vector in modal coordinates
$\{q\}$	Vector of node displacements normal to the surface of structure
$\{q\}_s$	Nodal displacement vector
$[\overline{q_j q_k}]$	Deflection covariance matrix

$[S]_1, [S]_2$	Strain - deflection transformation matrices
$[S_F(\Omega)]$	Cross spectral density matrix of $\{F\}$
$S_P(\Omega)$	Spectral density function of P
$\{\delta\}$	Element nodal displacement vector
$\{\epsilon\}$	Element strain vector
$[\overline{\epsilon_r \epsilon_s}]$	Strain covariance matrix
ζ	Damping ration
λ, μ	Proportionality constant between damping and stiffness and inertia, respectively
$\{\xi\}$	Amplitude vector in modal coordinates
$\{\phi^{(j)}\}$	j -th normalized eigen vector
$[\Phi]$	Modal transformation matrix
ω	Linear undamped frequency
ω_{eq}	Equivalent linear frequency
Ω	Angular frequency

I. INTRODUCTION

The response of outside surfaces or skins of an aircraft structure to high intensity acoustic pressure levels has been the subject of considerable research effort (References 1 and 2, for example). This complex problem may separate to three parts:

- (1) Prediction of the acoustic loading,
- (2) Determination of the response of complex structural panels to this excitation, and
- (3) Estimation of the fatigue life.

There are considerable data available to predict the acoustic loads on an aircraft structure due to the many possible sources of high sound levels. These sources are normally classified as to propulsion system noise sources and aerodynamic noise sources. Noise prediction methods for various sources have been summarized in two excellent reports (References 2 and 3).

Basic design considerations and procedures to estimate the fatigue life, various cumulative damage theories, and fatigue curves describing the S-N characteristics for various materials are given in References 1 and 2. Fatigue design data for bonded aluminum structures can be found in Reference 4.

The present study concentrates on the response aspects of the problem, even there is considerable overlapping of these three areas. Special effort is given to incorporate the nonlinear response effect, which is mainly due to large deflections, into the analysis. The

reason for this is simple because there is enormous discrepancy between the measured and the computed responses. Those computed results were based on linear or small deflection theory while test panels responded nonlinearly with large deflection at sufficiently high intensity noise levels. These documented evidence are summarized briefly in Section II.

In approaching this problem of random excitation of nonlinear systems, one can turn to a number of prior investigations. Therefore, an important task of this study is thus to review, classify, and evaluate these existing techniques. The evaluation is based on some realistic considerations from the point view of their application to complex panels of aircraft structure. These are presented in Section III. And it becomes evident from the survey that more research is needed in the nonlinear random vibration area.

Section IV gives a mathematical formulation, which is based on the finite element method and the equivalent linearization approach, for complex structures subjected to random acoustic excitation. An iterative scheme used in the solution procedure and a simplified flow chart are also presented. Ratio of nonlinear root mean-square (RMS) deflection to linear RMS deflection, that can be expressed in terms of linear frequency, equivalent linear frequency, and power spectral density (PSD) of the excitation, is given. A very crude estimate on the improvement of predicting responses using large deflection theory is also included in this section.

Section V gives a review of the advances in the development of geometrical stiffness matrices for various finite elements. These matrices are required in the nonlinear random vibration analysis.

A "Quasi-Linearization" method which is, at present, in the conceptual phase is discussed briefly in Section VI. Conclusions and recommended future work are given in Section VII.

II. EVIDENCE OF LARGE AMPLITUDE NONLINEARITY

Some aspects of current knowledge about the response of structural panels to high intensity noise are discussed in this section. Experiments and tests on aircraft structural components have displayed behavior that is not consistent with linear or small deflection theory assumptions. The deviations, which differ for various structural configurations, are suggestive of two types of nonlinearity sources:

- (1) Nonlinear damping ratio, and
- (2) Nonlinear relation between displacement and strain.

Commonly used methods for determining damping ratio are the bandwidth method by measuring half-power widths at modal resonances, and the decay rate method by measuring the logarithmic decrement on decaying modal response traces. The values of damping ratio range generally from 0.001 to 0.025 for common panel constructions used in aircraft structure. For such relatively small damping coefficients, it is a reasonable assumption that the effects to the structural behavior due to nonlinear damping would be also small in comparison with due to large amplitude. This has been an observed evidence in many experiments and is discussed and summarized in the following.

Many documents, for example references 5 to 8, have repeatedly reported poor comparison between measured and calculated RMS responses. They all observed that the test panels responded with large deflections at high sound pressure levels, and the computed responses were based on linear small deflection theory. This is the major reason

that causes the huge discrepancy between measured and calculated results. This experimental evidence was taken from:

(1) ASD-TDR-62-26 (Reference 5)

Fitch et al observed the nonlinear stress response at relatively low siren excitation level for conventional skin-stringer panels as shown in Figure 1. Their explanation to this nonlinear behavior was the diaphragm action which limits the amplitude of deflection of the vibrating plates. Examination of Figure 1 indicates that as the excitation level is raised above 104 dB, the structural response begins deviating away from the linear assumption.

(2) AFFDL-TR-68-44 (Reference 6)

Three types of typical aerospace structures were tested at an average sound pressure level (SPL) of 157 dB. They are a skin-stringer 3-bay, a honeycomb sandwich, and a corrugated sandwich panel. Measured RMS deflection on the skin-stringer panel is surprisingly large of 0.064 inch, or twice the skin thickness. Measured honeycomb panel RMS deflection is 0.031 inch, or 1.4 thickness of the top skin. Comparisons of RMS deflection showed that calculated skin-stringer panel deflection is 34-percent high and calculated honeycomb panel deflection is 5-percent high only. Comparisons between calculated and measured RMS stresses were poor. They are shown in Table 1 taken from Reference 6.

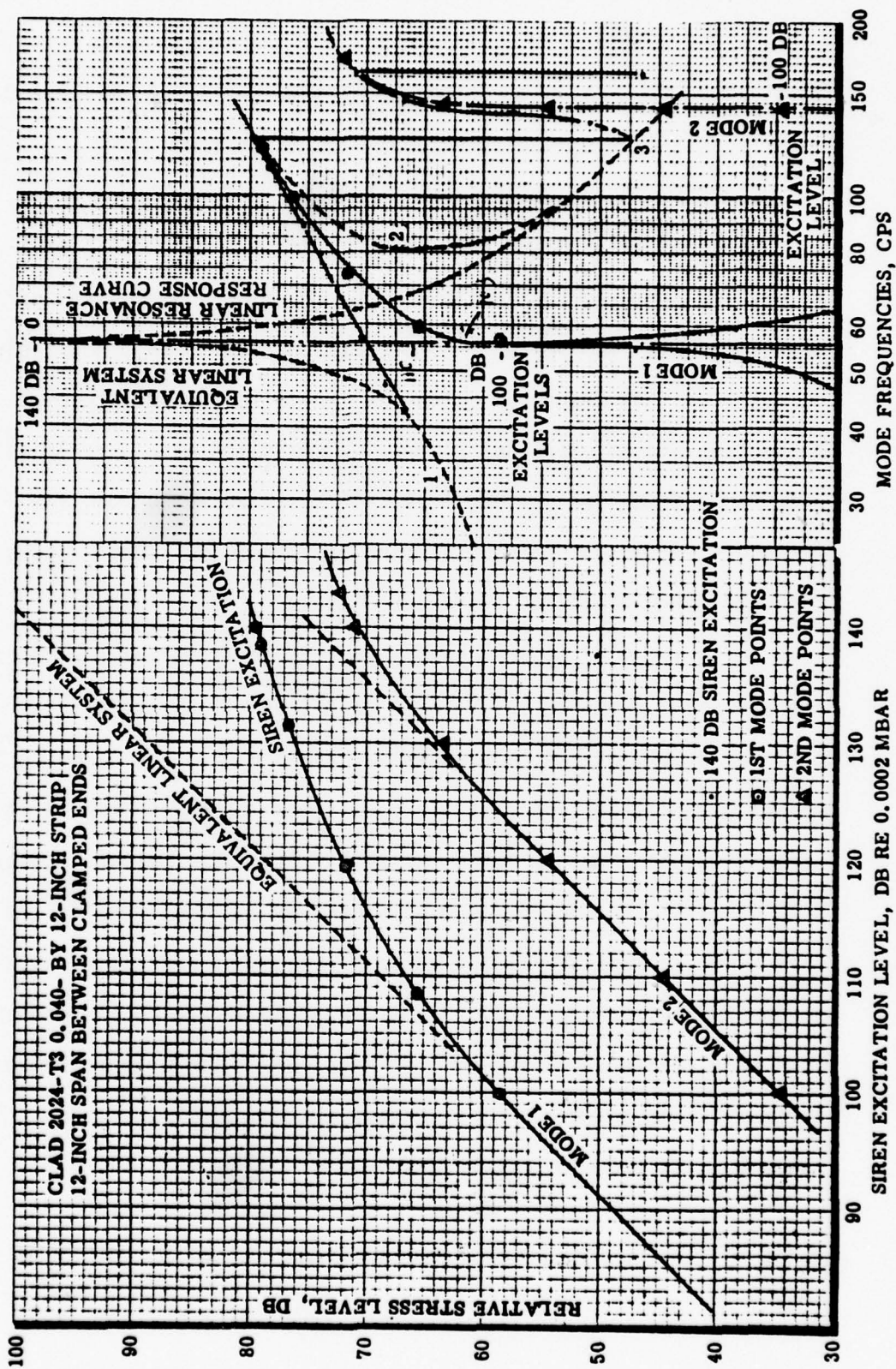


Figure 1. Nonlinear Stress Response

TABLE 1 - STRESS COMPARISON

Panel type	Stress component	RMS stress (kpsi)						
		Calculated	Measured on panel:					Measured average
			A	B	C	D	E	
Skin-stringer	$\sqrt{\sigma_x^2}$	7.7	2.2	2.9	2.5	---	2.2	2.5
	$\sqrt{\sigma_y^2}$	2.4	0.63	0.94	0.78	1.1	0.84	0.87
Honeycomb	$\sqrt{\sigma_x^2}$	2.6	2.0	1.8	1.2	1.2	1.3	1.5
	$\sqrt{\sigma_y^2}$	2.1	1.3	1.3	0.91	0.84	1.3	1.1

Strain components were also measured with rosette strain gauges mounted on both the upper and lower surfaces. Appreciable membrane stress was recorded, this implies the panels responding with large deflections. The stress levels measured were actually very low, showing that the nonlinearity is not associated with yield or plasticity of the material, but rather with coupling of inplane and out-of-plane displacements.

(3) AFFDL-TR-71-126 (Reference 7)

Three nine-bay, cross-stiffened, graphite-epoxy panels were exposed in a broad-band 166 dB SPL acoustic environment. Under loud-speaker excitation, the lowest natural frequency was at 174 Hz. But under the acoustic pressure in the progressive wave test chamber, the lowest frequency increased from 200 Hz to 290 Hz as the SPL was increased from 139 dB to 166 dB. The increase in natural frequency with increasing pressure level is attributed to the large deflections

of the panel response. Again strain comparisons were poor as shown in Table 2 taken from Reference 7. And deflection was not measured.

TABLE 2 - COMPARISON OF EXPERIMENTAL RESULTS
OF CROSS-STIFFENED PANELS WITH RESULTS
USING UNSTIFFENED PLATE THEORY

Approach	Method	Fundamental Frequency	Strain at $x = 0, y = \frac{b}{2}$	Strain at $x = \frac{a}{2}, y = 0$
		(Hz)	(micro-inch/inch-rms)	(micro-inch/inch-rms)
Analytic	Simplified theory (Beam Functions)	182	180 ⁽²⁾	406 ⁽²⁾
Analytic	Finite Element (REDYN)	180	180 ⁽²⁾	348 ⁽²⁾
Experi- mental	Test Panel A-GG-B-2	187 ⁽¹⁾	96 ⁽⁴⁾	160 ⁽³⁾
Experi- mental	Test Panel A-GG-B-3	170 ⁽¹⁾	74 ⁽⁴⁾	164 ⁽³⁾
<p>(1) Obtained during damping factor determination under loudspeaker excitation</p> <p>(2) Strain response to fully correlated, white noise excitation of 1.2×10^{-6} psi²/Hz</p> <p>(3) Strain gage No. 2 reading during 139 db run</p> <p>(4) Strain gage No. 7 reading during 139 db run</p>				

(4) AFFDL-TR-77-45 (Reference 8)

Total of ten bonded aluminum panels were tested. Table 3 taken from Reference 8 gives the ratio of the fundamental frequency obtained under the loud-speaker excitation to the frequency at which the peak strain PSD occurred in the sonic fatigue test at 166 dB overall SPL. Again, the agreement between the test strains and the predicted strain was poor, as shown in Table 4 (taken from Reference 8) and no measured deflection was reported.

Therefore, a conclusion can be reached that any real structure designed for service in high sonic environment regions would respond with large deflection nonlinearity.

TABLE 3 - FREQUENCY AND DAMPING DATA

PANEL	DAMPING FACTOR ⁽¹⁾	NATURAL FREQUENCIES				⁽²⁾ f_{166}	$\frac{f_{166}}{f_1}$
		$f_1 \equiv$ FIRST MODE	$f_2 \equiv$ SECOND MODE	$f_3 \equiv$ THIRD MODE	$f_4 \equiv$ FOURTH MODE		
		(Hz)	(Hz)	(Hz)	(Hz)	(Hz)	
A-1-1	0.018	137	189	265	340	218	1.59
A-1-2	0.015	97	147	179	266	204	2.10
A-2-1	0.011	144	209	284	386	212	1.47
A-2-2	0.009	141	203	279	376	211	1.50
A-3-1	0.012	165	245	-	-	205	1.21
A-3-2	0.011	141	210	295	413	207	1.47
A-4-1	0.023	80	114	155	226	140	1.75
A-4-2	0.009	84	124	172	235	144	1.71
A-5-1	0.012	103	155	211	292	150	1.46
A-5-2	0.014	99	153	215	300	143	1.44
Average	0.0134						

(1) Nondimensional viscous damping factor, $\frac{c}{c}$

(2) The parameter f_{166} was the frequency of the predominant strain response at 166 dB overall SPL.

TABLE 4 - STRAIN PREDICTIONS BASED ON THE ASSUMPTION
OF FULLY CLAMPED EDGES OF A PLATE

TEST PANEL BEING SIM- ULATED	OVER- ALL SPL	PRESS- URE PSD	RMS STRAINS					
			GAGE NO. 2		GAGE NO. 4		GAGE NO. 11	
			TEST	PRE- DICTED	TEST	Pre- DICTED	TEST	Pre- DICTED
	(dB)	(psi ² /Hz)	(μ"/")	(μ"/")	(μ"/")	(μ"/")	(μ"/")	(μ"/")
A-2-1	142	5.0×10^{-6}	32	445	27	111	84	270
A-4-1	145	3.4×10^{-5}	75	766	110	192	178	466

III. REVIEW OF EXISTING APPROACHES ON NONLINEAR RANDOM VIBRATION

Methods used to model structural systems can be basically divided into analytical methods and numerical methods. Analytical approaches are usually focused at obtaining explicit closed-form quantitative results for simple structural configurations. For complex structures such as found in aircraft design, numerical methods must invariably be employed to accurately model the complex configuration. However, analytical methods can generally provide more insight and lead to a better understanding of the problem. And other approximate and numerical techniques can then be extended from there. In this section, both analytical and numerical approaches for solving random excitation of nonlinear systems are reviewed.

1. Analytical Methods

a. Fokker-Planck Equation Approach. The most general extension of the Fokker-Planck equation method to the multiple-degree-of-freedom (multiple-DOF) systems of nonlinear second order equations is due to Caughey (References 9 and 10). One great advantage of this method over all of the other approaches is that it gives an exact solution. However, this should not be conceived that all problems relating to the response of nonlinear systems with random excitation have been solved. By the fact, exact solutions of the steady-state probability function have only been found for certain restricted classes of problems provided:

- (1) The only energy dissipation in the system arises from damping forces which are proportional to the velocity,
- (2) The exciting forces are uncorrelated Gaussian white noise,
- (3) The spectral density matrix of the excitation is proportional to the damping matrix of the system, and
- (4) The restoring force vector of the system is derivable from a potential.

The solution of the time-independent Fokker-Planck equation under these conditions represents a very significant accomplishment. Problems of simple structures which satisfy these four conditions were solved by Herbert (Reference 11 and 12). Yet, many problems of practical interest do not satisfy those conditions necessary for a solution. The required relationship between the excitation and the damping matrix is particularly restrictive. In addition, the transitional probability density function generally can not be found with the Fokker-Planck approach. Without this transitional probability, it is generally impossible to obtain the correlation function and PSD of the response. Thus, a number of approximate techniques have been developed to treat a broader class of problems than is presently possible with the exact analysis. These are the equivalent linearization technique, perturbation method, etc.

b. Equivalent Linearization Approach. This method was originated by Krylov and Bogoliubov (Reference 13). Caughey (Reference 14) has extended the equivalent linearization technique to systems of non-

linear differential equations. In his formulation, the correlation function matrix of the excitation must be diagonalized by the same transformation that diagonalizes the linear mass, damping, and stiffness matrices. This represents a rather severe limitation and in particular precludes the application of this formulation to dynamic systems which are excited randomly at only several nodal points. Lin (Reference 15) used the equivalent linearization method with a single-mode approach and obtained response for a rectangular panel subjected to randomly-varying loadings. Seide (Reference 16) has employed the formulation by Caughey and obtained solution for a simple beam subjected to uniform pressure excitation uncorrelated in time.

Foster (Reference 17), and Iwan and Yang (Reference 18) have extended the equivalent linearization technique by removing the restriction imposed on the transformation. Foster's formulation is much general. He first replaced the original n -DOF second order system by a $2n$ -DOF first order system. The determination of the equivalent linear stiffness coefficients is then accomplished by the inversion of a $2n \times 2n$ mean-square matrix and an iterative procedure. Practical application of this approach to a simple deep-ocean tower frame structure is reported in Reference 19.

c. Perturbation Approach. A perturbation method, based on classical perturbation theory, was developed by Crandall (Reference 20) to obtain approximate solutions to nonlinear systems, containing a small parameter, excited by a weakly stationary random Gaussian processes. In principle, the perturbation approach can be extended

to systems of coupled nonlinear equations in which the nonlinearities contain a small parameter. Lyon (Reference 21) used this method to study the responses of a nonlinear strong. Tung, Penzien, and Horonjeff (Reference 22) used the perturbation procedure to a two DOF system. For complex structures, however, the algebraic operations may become so unwieldy that the method is no longer practical. In addition, that there are certain subtle questions about the convergence of the power-series expansion for the nonlinear response still remaining unanswered.

d. Other Approximate Methods. Fox et al (Reference 23) have developed three new approaches: (1) Direct Evaluation of Spectra, (2) Estimates of Equilibrium Distribution, and (3) Generalized Kinetic Equation. They have applied the direct evaluation of spectra method to a hinged uniform beam with the assumptions that the force spectrum is white and that all DOF are equally forced and have the same damping factor. Extension of these methods to complex structures would certainly require considerable efforts.

2. Numerical Methods

Numerical methods can be subdivided into two categories: numerical solutions to differential equations or finite difference method, and matrix displacement method based on discrete element idealization or finite element method.

a. Finite Difference Approach. Numerical solutions to differential equations are somewhat restricted so that these techniques can

be practically applied only to simple structural configurations. Belz (Reference 24) used the finite difference approach and obtained statistical response for a simple uniform beam subjected to a single concentrated load at the midspan of the beam.

b. Finite Element Method. Application of the finite element methods to linear structures subjected to random excitations have been presented, for example, by Jacobs and Lagerquist (References 6 and 25), Olson and Lindberg (Reference 26), Jacobson (Reference 7), and Olson (Reference 27). In Reference 26, refinement on the continuity between the stiffeners and the panel itself was introduced. Olson in Reference 27 presented a consistent formulation for the cross spectral density matrix of the excitation. Both should give improvement in the accuracy of predicting random responses for linear structures.

Application of the finite element method to deep-ocean towers has been given by Foster (Reference 19) and to off-shore towers by Penzien et al (Reference 28). In the problems they solved, the nonlinear effects can be expressed explicitly in terms of the displacements or velocities. This is not the case for problems of complex panel responses to high intensity noise. The nonlinear effect due to large deflections or the geometrical stiffness matrix is not known a priori.

Extension of the finite element approach to complex structures under high noise environment with large deflection nonlinear effect is the purpose of this research. A matrix formulation which is based on the finite element method and the equivalent linearization technique is developed and presented in next section.

IV. MATHEMATICAL FORMULATION

1. FINITE ELEMENT REPRESENTATION

The structures considered will be restricted to stable structural systems which are highly resonant, that is, with little damping. These are typical properties for panel components of aircraft structure.

In the matrix structural analysis, structure is idealized into a finite number of discrete structural elements connected at node points. The physical properties of the structure are assumed to be lumped into individual elements. The stiffness equations of motion for such an element under the influence of dynamic loading, inertia, damping, elastic, and nonlinear large deflection characteristics are:

$$[m]\{\ddot{\delta}\} + [c]\{\dot{\delta}\} + ([k] + [k^g(\{\delta\})])\{\delta\} = \{f(x)\} \quad (1)$$

Where $\{\delta\}$ and $\{f\}$ are vectors of nodal displacements and applied forces, respectively. The consistent mass $[m]$, damping $[c]$, and linear stiffness $[k]$ matrices have been developed for almost every beam, plate, and shell elements available. The element geometrical stiffness matrix or nonlinear stiffness matrix $[k^g(\{\delta\})]$, which is displacement dependent and is induced due to large deflections, will be discussed in Section V.

By assembling all the elements, and applying the kinematic boundary conditions, the equations of motion of the structure are:

$$[M]_s\{\ddot{q}\}_s + [C]_s\{\dot{q}\}_s + ([K]_s + [K^g(\{q\}_s)])\{q\}_s = \{F(x)\}_s \quad (2)$$

In which $\{q\}_s$ and $\{F\}_s$ denote the vectors of nodal displacements and forces of the structure. Matrices $[M]_s$, $[C]_s$, $[K]_s$, and $[K^g]_s$ are the mass, damping, stiffness, and geometrical stiffness coefficients of the structure, respectively.

Applying the static condensation (or Guyan reduction) and retaining only those DOF, say m of them, normal to the surface of the structure, Eq. (2) may be written as:

$$\underset{m \times m}{[M]}\{\ddot{q}\} + \underset{m \times m}{[C]}\{\dot{q}\} + (\underset{m \times m}{[K]} + \underset{m \times 1}{[K^g]}\{\bar{q}\})\{q\} = \underset{m \times 1}{\{F(t)\}} \quad (3)$$

where $\{q\}$ is a vector containing all nodal deflections normal to the panel structure. The matrices $[M]$, $[C]$, $[K]$, and $[K^g]$ denote the reduced mass, damping, stiffness, and geometrical stiffness, respectively.

2. DAMPING REPRESENTATION

For certain forms of damping, the coupled nonlinear equations of motion, Eq. (3), can be reduced to a set of equations which contain coupling only in the nonlinear terms. This requires the determination of the eigen values and eigen vectors of the undamped linear system

$$\omega_j^2 [M]\{\phi^{(j)}\} = [K]\{\phi^{(j)}\} \quad j = 1, 2, \dots, m \quad (4)$$

in which ω_j is the natural frequency and $\{\phi^{(j)}\}$ is the corresponding j -th mode shape of the linear structure.

Apply a coordinate transformation, from the nodal displacements to the modal coordinates, by

$$\underset{m \times 1}{\{q\}} = [\underset{m \times n}{\{\phi^{(1)}\}}, \underset{n \times 1}{\{\phi^{(2)}\}}, \dots, \underset{n \times 1}{\{\phi^{(n)}\}}] \underset{n \times 1}{\{\xi\}} = \underset{m \times n}{[\Phi]} \underset{n \times 1}{\{\xi\}} \quad n \leq m \quad (5)$$

in which $\{\xi\}$ represents a vector of modal coordinates. Substituting Eq. (5) into Eq. (3) and premultiplying by the transpose of $[\Phi]$, Eq. (3) becomes

$$[M]\{\ddot{\xi}\} + [\Phi]^T[C][\Phi]\{\dot{\xi}\} + [K]\{\xi\} + [\Phi]^T[K^g(\xi)][\Phi]\{\xi\} = \{P(x)\} \quad (6)$$

where $\{P\} = [\Phi]^T\{F\}$ is the generalized force vector in modal coordinates. The terms M_j and K_j are the j -th generalized mass and stiffness defined by

$$\begin{aligned} M_j &= \{\phi^{(j)}\}^T [M] \{\phi^{(j)}\} \\ K_j &= \{\phi^{(j)}\}^T [K] \{\phi^{(j)}\} = \omega_j^2 M_j \end{aligned} \quad j = 1, 2, \dots, n \quad (7)$$

The equations of motion, Eq. (6) will contain coupling only in the nonlinear terms if the viscous damping (Reference 29) is proportional to inertia, stiffness, or both, that is

$$[C] = \mu[M] + \lambda[K] \quad (8)$$

where μ and λ are proportionality constants. Then, the j -th generalized damping coefficient is given by

$$\begin{aligned} C_j &= \{\phi^{(j)}\}^T [C] \{\phi^{(j)}\} \\ &= \mu M_j + \lambda K_j \end{aligned} \quad j = 1, 2, \dots, n \quad (9)$$

Structural damping is another form of damping that allows it to be uncoupled, that is

$$[C] = i g [K] \quad (10)$$

where $g(g \ll 1)$ is the structural damping coefficient. The j -th generalized damping coefficient is

$$C_j = i g K_j \quad (11)$$

Sometimes it is more convenient to represent the damping as a fraction of critical damping. The modal damping ratio ζ_j represents the fraction of critical damping in the j -th mode. This ratio is related to viscous damping proportionality constants μ and λ in Eq. (8), and to structural damping coefficient g in Eq. (10) by the relationships

$$2\zeta_j = \begin{cases} \frac{\mu}{\omega_j} + \lambda \omega_j, & \text{for viscous damping} \\ g, & \text{for structural damping} \end{cases} \quad (12)$$

When damping can be represented by proportional viscous or structural damping, then equations of motion, Eq. (6), can be written as

$$[M]\{\ddot{\epsilon}\} + [C]\{\dot{\epsilon}\} + [K]\{\epsilon\} + [\Phi]^T [K^g(\epsilon)] [\Phi]\{\epsilon\} = \{P\} \quad (13)$$

The j -th row of Eq. (13) has the form

$$\begin{aligned} M_j \ddot{\epsilon}_j + C_j \dot{\epsilon}_j + K_j \epsilon_j + \sum_{k=1}^n \{\phi^{(j)}\}^T [K^g] \{\phi^{(k)}\} \epsilon_k &= P_j \\ \text{or, } M_j \ddot{\epsilon}_j + C_j \dot{\epsilon}_j + K_j \epsilon_j + \sum_{k=1}^n \epsilon_k \sum_{r=1}^m \sum_{s=1}^m \Phi_{rj} K_{rs}^g \Phi_{sk} &= P_j \end{aligned} \quad (14)$$

which has coupling only in the nonlinear term.

3. EQUIVALENT LINEARIZATION APPROACH

The basic idea of the equivalent linearization method (References 14 and 17) is to replace the actual system, Eq. (13) or (14), with a set of equations of the form

$$M_j \ddot{\xi}_j + C_j \dot{\xi}_j + K_j^{eq} \xi_j + e_j(\xi_1, \xi_2, \dots, \xi_n) = P_j$$

$$j = 1, 2, \dots, n \quad (15)$$

where K_j^{eq} is an equivalent linear stiffness constant, and e_j is the error of linearization or equation deficiency term.

If this error term e_j is neglected, then Eq. (15) is linear and it can be readily solved. The smaller that the error is, the smaller the error in neglecting it, and the better approximate solution to Eq. (14) will be obtained. To this end, the n equivalent linear stiffness constants, K_j^{eq} , are chosen in such a way that the mean-square, $E[e_j^2]$, is minimized. The error of linearization is

$$e_j = K_j \xi_j - K_j^{eq} \xi_j + \sum_{k=1}^n \{\phi^{(j)}\}^T [K^g(\xi)] \{\phi^{(k)}\} \xi_k$$

$$j = 1, 2, \dots, n \quad (16)$$

which is the difference between Eq. (14) and Eq. (15). From Eq. (16) it is apparent that e_j depends upon the equivalent linear stiffness constant K_j^{eq} . It is these constants which will vary in order to minimize the n values of $E[e_j^2]$ requiring the following equation for K_j^{eq}

$$\frac{\partial E[e_j^2]}{\partial K_j^{eq}} = 0$$

$$j = 1, 2, \dots, n \quad (17)$$

where the operator $E[]$ denotes the statistical average or mathematical expectation of the appropriate variables. Substituting Eq. (16) into Eq. (17) and interchanging the order of differentiation and expectation, condition of Eq. (17) reduces to

$$K_j E[\xi_j^2] - K_j^{eq} E[\xi_j^2] + E[\xi_j \sum_{k=1}^n \{\phi^{(j)}\}^T [K^g(\xi)] \{\phi^{(k)}\} \xi_k] = 0$$

$j = 1, 2, \dots, n$

(18)

Solving for the equivalent linear stiffness constant K_j^{eq} , Eq. (18) gives

$$K_j^{eq} = K_j + \frac{E[\xi_j \sum_{k=1}^n \{\phi^{(j)}\}^T [K^g(\xi)] \{\phi^{(k)}\} \xi_k]}{E[\xi_j^2]}$$

$j = 1, 2, \dots, n$

(19)

Note that Eq. (19) is not an explicit equation for K_j^{eq} , since the expectations appearing on the right-hand side depend on K_j^{eq} . In addition, the geometrical stiffness matrix which is needed in the numerator is not known a priori, and the coupling of the modal displacements causes some difficulties in evaluating the expectation. Therefore, one has turn to simpler approach with a single-mode approximation solution. Further study is needed to evaluate Eq. (19) numerically or by some other means.

4. SINGLE-MODE APPROACH AND MEAN-SQUARE RESPONSES

For most of the sonic fatigue analyses in practice, only a single-mode approach and linear small deflection theory is commonly employed. The inclusion of the large deflection into the analysis with a single-mode approximation represents a significant improvement of the design tools for complex structural panels.

Thus, if a single-mode approximation (usually the fundamental mode) is assumed, and also if the excitation pressure to the structure is stationary, is Gaussian, and has a zero mean. The expectations in Eq. (19) can be evaluated, and the equivalent linear stiffness constant becomes

$$K^{eq} = K + 3\{\phi\}^T [K^g] \{\phi\} E[\xi^2] \quad (20)$$

in which the subscript j has been dropped. Note again that Eq. (20) is not explicit for K^{eq} , since the expectation $E[\xi^2]$ depends on K^{eq} , and also the geometrical stiffness matrix

which is displacement dependent is not known a priori. Therefore, an iterative scheme is introduced to determine K^{eq} from Eq. (20). This will be presented later. At present, let us assume that a satisfactory equivalent linear stiffness constant has been found.

By dropping the error of linearization $e(\xi)$ from Eq. (15), the single-mode approximation solution of Eq. (3) is obtained from

$$M\ddot{\xi} + C\dot{\xi} + K^{eq}\xi = P \quad (21)$$

$$\text{or} \quad \ddot{\xi} + 2\zeta\omega\dot{\xi} + \omega_{eq}^2\xi = \frac{P}{M} \quad (22)$$

where $\omega_{eq} = (K^{eq}/M)^{1/2}$ is an equivalent linear frequency, and ζ is the modal damping ratio related to the linear frequency ω by Eq. (12).

Now a random analysis of the modal equation (22) may be easily carried out to yield the PSD for the modal amplitude ξ as

$$S_{\xi}(\Omega) = S_P(\Omega) |H(\Omega)|^2$$

$$= \{\phi\}^T [S_F(\Omega)] \{\phi\} |H(\Omega)|^2 \quad (23)$$

in which $[S_F(\Omega)]$ is the cross spectral density matrix of the noise excitation $\{F\}$, and the frequency function $H(\Omega)$ is given by

$$H(\Omega) = \frac{1}{M(\omega_{eq}^2 - \Omega^2 + i 2\zeta \omega \Omega)} \quad (24)$$

The mean-square response of modal amplitude is related to $S_\xi(\Omega)$ by

$$\begin{aligned} E[\xi^2] &= \int_0^\infty S_\xi(\Omega) d\Omega \\ &= \int_0^\infty \frac{S_p(\Omega) d\Omega}{M^2 [(\omega_{eq}^2 - \Omega^2)^2 + (2\zeta \omega \Omega)^2]} \end{aligned} \quad (25)$$

For lightly damped ($\zeta \leq 0.05$) structures, the response curves will be highly peaked at ω_{eq} . The integration of Eq. (25) can be greatly simplified if $S_p(\Omega)$ or $[S_F(\Omega)]$ can be considered to be constant in the frequency band surrounding the resonance peak, so that

$$\begin{aligned} E[\xi^2] &= S_p(\omega_{eq}) \int_0^\infty |H(\Omega)|^2 d\Omega \\ &= \frac{\pi}{4M^2 \omega \omega_{eq}^2 \zeta} \{\phi\}^T [S_F(\omega_{eq})] \{\phi\} \end{aligned} \quad (26)$$

The covariance matrix of the nodal deflections $\{q\}$ can be obtained by use of the coordinate transformation given in Eq. (15), then

$$[\overline{q_j q_k}] = \{\phi\} E[\xi^2] \{\phi\}^T \quad j, k = 1, 2, \dots, m \quad (27)$$

The diagonal elements of $[\overline{q_j q_k}]$ are the mean-square values of the nodal deflections, and the off-diagonal terms are time averages of products of deflections at different nodes. The mean-square node deflection is simply

$$\begin{aligned} \overline{q_j^2} &= E[\xi^2] \phi_j^2 \\ &= \frac{\pi \phi_j^2}{4 M^2 \omega \omega_{eq}^2 \zeta} \{\phi\}^T [S_F(\omega_{eq})] \{\phi\} \quad j=1,2,\dots,m \quad (28) \end{aligned}$$

in which ϕ_j is the j -th element in modal vector $\{\phi\}$.

The element strains and nodal deflections are related by

$$\{\epsilon\} = [S]_1 \{q\} + [S]_2 \{q\} \quad (29)$$

in which $\{\epsilon\}$ is the vector of element strains, $[S]_1$ and $[S]_2$ are the strain-deflection transformation matrices. $[S]_1$ is the usual transformation matrix based on linear theory, and $[S]_2$ represents the inplane strains due to large deflections. Eq. (29) is based on an appropriate linearization of the nonlinear strain-displacement relations (Reference 30) which will be discussed in Section V. Thus, the strain covariance matrix is given by

$$\begin{aligned} [\overline{\epsilon_r \epsilon_s}] &= [S]_1 [\overline{q_j q_k}] [S]_1^T + [S]_1 [\overline{q_j q_k}] [S]_2^T \\ &\quad + [S]_2 [\overline{q_j q_k}] [S]_1^T + [S]_2 [\overline{q_j q_k}] [S]_2^T \quad (30) \end{aligned}$$

The diagonal elements are the mean-square values of element strains.

5. INTERATION PROCEDURE AND FLOW CHART

As it was pointed earlier in Eq. (20) that the equivalent linear stiffness constant K^{eq} depends upon its response, which, inturn, depends upon K^{eq} . And also, the geometrical stiffness matrix $[K^g]$ is response dependent and it is not known a priori. In this way an interactive approach to final solution occurs. It makes no difference at which point in the interation cycle the process begins. One certainly can assume either equivalent linear stiffness or responses at the outset. The process will converge to an answer provided the structure considered is a stable one. And all panel structures of aircraft or space vehicle under sound environment are examples of such stable types.

Suppose, for definiteness, one desires to estimate the initial equivalent linear stiffness constant as K_1^{eq} . The fact that both $E[\xi^2]$ and $[K^g]$ can be approximated using the solutions of the linear equations of motion (in Eq. (26), ω_{eq} has to be replaced by ω), facilitates the initial estimate of K^{eq} through Eq. (20) as

$$K_1^{eq} = K + 3\{\phi\}^T [K^g]_0 \{\phi\} E[\xi^2]_0 \quad (31)$$

This calculated initial estimate of K_1^{eq} can be used to obtain refined estimate of $E[\xi^2]_1$ and $[K^g]_1$, that implies K_2^{eq} through Eq. (20) in the same way that Eq. (31) implied K_1^{eq} . As the interactive process converges on the j-th cycle, the relation

$$\begin{aligned} K_j^{eq} &= K + 3\{\phi\}^T [K^g]_{j-1} \{\phi\} E[\xi^2]_{j-1} \\ &\approx K_{j-1}^{eq} \end{aligned} \quad (32)$$

becomes satisfied. The number of cycles required to attain convergence depends on the nonlinear characteristics of the structure $[K^g]$, the intensity of the excitation $[S_F(\omega_{eq})]$, and the accuracy desired. The solution procedure is illustrated by a simplified flow chart shown in Figure 2.

6. ESTIMATION OF IMPROVEMENT

It is a very difficult task to give an accurate quantitative estimate on the improvement that would be made in predicting of random responses using the nonlinear formulation without developing the computer program and analyzing these problems given in Section II. A very crude estimate of root-mean-square (RMS) deflections, however, is possible. The ratio of RMS deflection based on large deflection nonlinear theory to RMS deflection using linear theory is given by

$$\frac{(\text{RMS deflection})_{NL}}{(\text{RMS deflection})_L} = \frac{\omega}{\omega_{eq}} \sqrt{\frac{S_P(\omega_{eq})}{S_P(\omega)}} \quad (33)$$

in which ω is the linear natural frequency, and ω_{eq} is the equivalent linear frequency. If the spectral density function of excitation $S_P(\omega)$ is a slowly varying function with respect to frequency, Eq. (33) can be further simplified to

$$\frac{(\text{RMS deflection})_{NL}}{(\text{RMS deflection})_L} \approx \frac{\omega}{\omega_{eq}} \quad (34)$$

Let us examine the data obtained in Reference 6 (see Section II (2)), because this was the only report that deflections were measured.

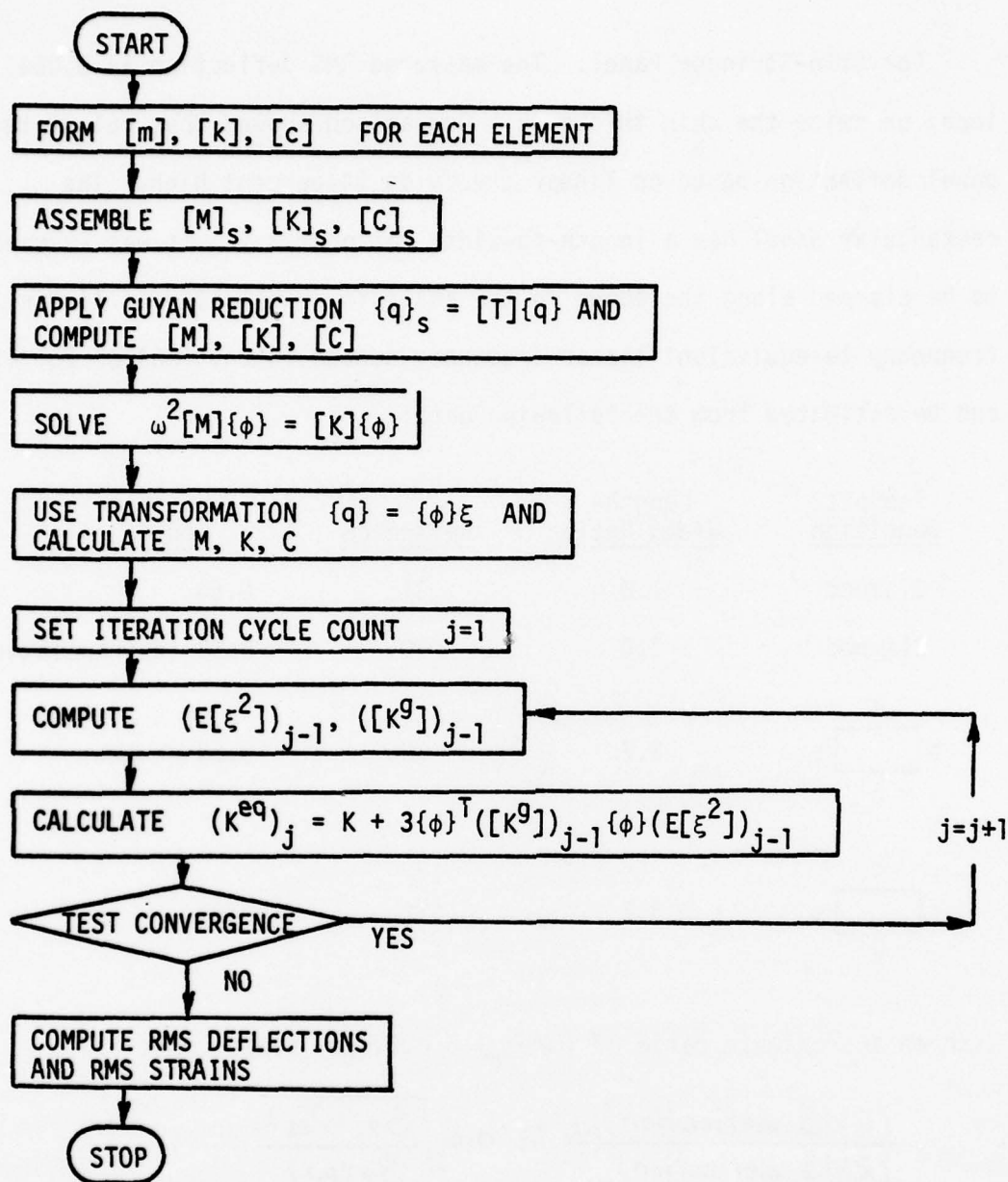
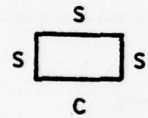
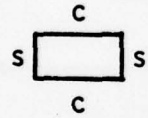


Figure 2. Simplified flow chart for complex panel nonlinear response to high intensity random loads.

For Skin-Stringer Panel. The measured RMS deflection is 0.064 inch, or twice the skin thickness. Comparison showed that calculated panel deflection based on linear theory is 34-percent high. The rectangular panel has a length-to-width ratio of 3.7. It was assumed to be clamped along the edges in the analysis. The ratio of linear frequency to equivalent linear frequency needed in Eq. (33) or Eq. (34) can be estimated from the following data:

<u>Support Condition</u>	<u>Length-Width Ratio</u>	<u>Reference</u>	<u>$\frac{\omega}{\omega_{eq}}$</u>
Clamped	1.0	31	0.64
Clamped	2.0	30	0.58 (extrapolated)
	3.7	32	0.53
	3.7	32	0.65

With an approximate ratio of $(\omega/\omega_{eq}) \cong 0.6$, Eq. (33) becomes

$$\frac{(RMS \text{ deflection})_{NL}}{(RMS \text{ deflection})_L} \cong 0.6 \sqrt{\frac{S_p(\omega_{eq})}{S_p(\omega)}} \quad (35)$$

The RMS deflection based on large deflection theory would be very close to the measured value.

Without analyzing the problem with computer program, it is very difficult to give a quantitative estimate on the improvement in predicting response strains using nonlinear theory. The difficulties

come from the transformation matrices $[S]_1$ and $[S]_2$ in Eq. (30).

Therefore, no attempt is given for pursuing it further.

V. GEOMETRICAL STIFFNESS MATRICES

The geometrical stiffness matrix $[k^g(\{\delta\})]$ needed in the formulation has been developed for various types of elements. They have been used successfully, in conjunction with the consistent mass and stiffness matrices, in problems of large amplitude vibrations of complex structures. A brief review of the advances in the development of geometrical stiffness matrices and the associated large amplitude vibration problems is given in this section.

Extension of the finite element method to large amplitude vibrations of beams and rectangular plates was first reported by Mei (References 33 and 34). The geometrical stiffness matrix formulation for a rectangular plate element in Reference 34 was based on a modified form of the Berger's hypothesis (Reference 35). Nonlinear frequencies, which were obtained for rectangular plates with various edge support conditions, agreed well with the approximate analytical solutions. Results for some boundary conditions in Reference 34 were obtained for the first time.

One important thing which has to be mentioned at this moment is that the nonlinear frequencies determined from the large amplitude vibrational analysis using a "Quasi-Linearization" technique have the very same physical meaning, comparing Eq. (14) with Eqs. (21) and (22), as the equivalent linear frequency, ω_{eq} , using equivalent linearization method. Therefore, to gain a better understanding of large amplitude vibrations of plate and shell structures will certainly be helpful to build a more solid background for problems of random vibrations of

complex nonlinear structures. This has been learned to be true in the linear case.

Recently, Rao and his colleagues presented a novel simplified formulation for large amplitude vibrations of beams (Reference 36), rectangular plates (References 30 and 37), and circular plates (Reference 38). Their formulation is based on an appropriate linearization of the nonlinear strain-displacement relations and an iterative scheme of Mei's (Reference 33) to obtain the nonlinear frequencies. This novel linearization technique was used in the strain-deflection relations of Eq. (29). Reddy and Stricklin (Reference 39) developed a linear and a quadratic isoparametric rectangular element to study large amplitude plate vibrations. Most recently, two triangular element formulations have been developed for large amplitude vibrations of thin plates of arbitrary planform. The first one (Reference 40) is consistent with the higher order bending element TRPLT1 (Reference 41) in NASTRAN program, and the second (Reference 41) is consistent with the high precision plate element of Cowper et al (Reference 43). Nonlinear frequencies obtained for numerical examples include rectangular, circular, rhombic, and isosceles triangular plates.

Raju and Rao (References 44 and 45) intended to develop a shell of revolution frustum for nonlinear vibration analysis of thin shells of revolution; however, their results failed to predict the "softening" type of nonlinear behavior as discussed by Evensen (Reference 46).

Geometrical stiffness matrices of various types of finite elements that have been developed for free vibration analysis involving large

deflection nonlinearities are listed in Table 5.

TABLE 5 - GEOMETRICAL STIFFNESS MATRICES
OF VARIOUS FINITE ELEMENTS

<u>Element Type</u>	<u>Reference</u>
Beam	Mei ³³ (1972) Rao et al. ³⁶ (1976)
Rectangular Plate	Mei ³⁴ (1973) Rao et al. ³⁰ (1976) Reddy and Stricklin ³⁹ (1977)
Rectangular (orthotropic) Plate	Rao et al. ³⁷ (1976)
Circular Ring (orthotropic) Plate	Rao et al. ³⁸ (1976)
Triangular Plate	Mei and Rogers ⁴⁰ (1977) Mei et al. ⁴² (1978)
Shell of Revolution	Raju and Rao ^{44,45} (1975,-76)

AD-A065 650

OHIO STATE UNIV RESEARCH FOUNDATION COLUMBUS
USAF-ASEE (1978) SUMMER FACULTY RESEARCH PROGRAM (WPAFB). VOLUM--ETC(U)
NOV 78 C D BAILEY

F44620-76-C-0052

UNCLASSIFIED

AFOSR-TR-79-0231

NL

2 OF 6

AD
A065650



VI. QUASI-LINEARIZATION METHOD

The same physical interpretation between the equivalent linear frequency and the nonlinear frequency, Eqs. (14), (21) and (22), leads to the idea that the method of quasi-linearization may also be applied to problems of complex nonlinear structures subjected to random loads. The quasi-linearization technique has been used successfully in predicting nonlinear frequencies for complex structures (References 33, 34, 36-40, 42, 44, 45). Nonlinear frequencies of higher modes can also be determined by this technique (References 30, 33, 40, 42). Those nonlinear frequencies, $\omega_{eq, j}$, can be employed to obtain an approximate random response of deflections from the relation

$$[\overline{q_r q_s}] = \sum_{j=1}^n \{\phi^{(j)}\} \frac{\pi \{\phi^{(j)}\}^T [S_F(\omega_{eq, j})] \{\phi^{(j)}\}}{4 M_j^2 \omega_j \omega_{eq, j}^2 \zeta_j} \{\phi^{(j)}\}^T \quad (35)$$

$r, s = 1, 2, \dots, m$

This method may very well be more promising than the equivalent linearization technique. However, both methods are worth pursuing further.

VII. CONCLUSIONS

1. SUMMARY OF RESULTS

The negligence of "large amplitude nonlinearity" in the analysis was identified to be the major factor that contributed to the enormous discrepancy between the measured test data and computed results. A mathematical formulation which is based on the finite element displacement method and the equivalent linearization technique is developed. Statistical responses of deflection and strain using a single-mode approximation can be expressed in terms of linear frequency, spectral density matrix of excitation, equivalent linear frequency, and transformation matrices. An iterative scheme which is used for determining the equivalent linear stiffness constant is presented. Advances in the development of geometrical stiffness matrices for various finite element is given. Finally, a concept of applying the quasi-linearization method to problems of nonlinear complex panel structures subjected to high intensity noise levels is discussed briefly.

2. RECOMMENDATIONS

a. Development of Computer Program. A computer program should be developed based on the finite element-equivalent linearization formulation presented. The computed results should be compared with the experiments.

b. Experiments. Carefully monitored and controlled experiments with simple structures (such as plates) and typical aircraft panels should be conducted at both low and high noise environment. Measurements

of deflection, strain, frequency, and pressure spectral density should be precisely recorded.

c. Refinement of Finite Elements. Refined finite element representations should be developed and incorporated into the computer program, such refinements, for example, as higher-order displacement function for beam element with various thin-walled cross-sections, variation in thickness for plate element, shallow shell element, anisotropic properties for modeling composite panels, etc.

d. Multi-Mode Approach. The improvement in response predictions using multiple-modes as to single-mode based on linear theory (low noise levels) should be investigated. Multi-modes must be employed in nonlinear analysis (high noise levels) if the improvement were found to be significant in the linear case.

e. Quasi-Linearization Method. More research study on this very promising approach should be conducted.

REFERENCES

1. Thomson, A.G.R. and Lambert, R.F., "Acoustic Fatigue Design Data", AGARD-AG-162-Part I and II, NATO Advisory Group for Aero. Res. and Dev., 1972.
2. Rudder, F.F., Jr. and Plumblee, H.E., Jr., "Sonic Fatigue Design Guide for Military Aircraft", AFFDL-TR-74-112, W-PAFB, OH, 1975.
3. Ungar, E.E. et al., "A Guide for Estimation of Aeroacoustic Loads on Flight Vehicle Surfaces", AFFDL-TR-76-91, W-PAFB, OH, 1977.
4. Jacobson, M.J., "Sonic Fatigue Design Data for Bonded Aluminum Aircraft Structures", AFFDL-TR-77-45, W-PAFB, OH, 1977.
5. Fitch, G.E. et al., "Establishment of the Approach to, and Development of Interim Design Criteria for Sonic Fatigue", ASD-TDR-62-26, W-PAFB, OH, 1962, pp. 42-46.
6. Jacobs, L.D. and Lagerquist, D.R., "Finite Element Analysis of Complex Panel to Random Loads", AFFDL-TR-68-44, W-PAFB, OH, 1968, pp. 55-59.
7. Jacobson, M.J., "Advanced Composite Joints; Design and Acoustic Fatigue Characteristics", AFFDL-TR-71-126, W-PAFB, OH, 1972, pp. 133 and 208.
8. Jacobson, M.J., "Sonic Fatigue Design Data for Bonded Aluminum Aircraft Structures", AFFDL-TR-77-45, W-PAFB, OH, 1977, pp. 54-57.
9. Caughey, T.K., "Derivation and Application of the Fokker-Planck Equation to Discrete Nonlinear Dynamic Systems Subjected to White Random Excitation", JASA, Vol. 35, Nov. 1973, pp. 1683-1692.
10. Caughey, T.K., "Nonlinear Theory of Random Vibrations", Advances in Applied Mechanics, Edited by Yih, C.S., Vol. 11, Academic Press, 1971, pp. 209-253.
11. Herbert, R.E., "Random Vibrations of a Nonlinear Elastic Beam", JASA, Vol. 36, Nov. 1964, pp. 2090-2094.
12. Herbert, R.E., "Random Vibrations of Plates with Large Amplitude", JAM, Vol. 32, Sept. 1965, pp. 547-552.

13. Krylov, N. and Bogoliubov, N., "Introduction to Nonlinear Mechanics", translated by Lefshetz, S. in Annals of Mathematical Studies, No. 11, Princeton University Press, N.J., 1943.
14. Caughey, T.K., "Equivalent Linearization Techniques", JASA, Vol. 35, 1963, pp. 1706-1711.
15. Lin, Y.K., "Response of a Nonlinear Flat Panel to Periodic and Randomly-Varying Loadings", J. Aerospace Sci., Sept. 1962, pp. 1029-1033, p. 1066.
16. Seide, P., "Nonlinear Stresses and Deflections of Beams Subjected to Random Time Dependent Uniform Pressure", ASME Paper No. 75-DET-23.
17. Foster, E.T., Jr., "Semilinear Random Vibrations in Discrete Systems", JAM, Vol. 35, Sept. 1968, pp. 560-564.
18. Iwan, W.D. and Yang, I.M., "Application of Statistical Linearization Technique to Nonlinear Multidegree-of-Freedom Systems", JAM, Vol. 39, June 1972, pp. 545-550.
19. Foster, E.T., Jr., "Predicting Wave Responses of Deep-Ocean Towers", Proceedings of the Conference on Civil Engineering in the Oceans, ASCE, 1968, pp. 75-98.
20. Crandall, S.H., "Perturbation Techniques for Random Vibration of Nonlinear Systems", JASA, Vol. 35, Nov. 1963, pp. 1700-1705.
21. Lyon, R.H., "Response of a Nonlinear String to Random Excitation", JASA, Vol. 32, August 1960, pp. 953-960.
22. Tung, C.C., Penzien, J. and Horonjeff, R., "The Effect of Runway Unevenness on the Dynamic Response of Supersonic Jet Transport", NASA CR-119, University of California, Berkeley, 1964.
23. Fox, H.L., Smith, P.W., Jr., Pyle, R.W. and Nayak, P.R., "Contributions to the Theory of Randomly Forced, Nonlinear, Multiple-Degree-of Freedom, Coupled Mechanical Systems", AFFDL-TR-72-45, W-PAFB, OH, 1973.
24. Belz, D.L., "A Numerical Study of Nonlinear Vibrations Induced By Correlated Random Loads", Int. J. Nonlinear Mechanics, Vol. 1, 1966, pp. 139-145.

25. Jacobs, L.D. and Lagerquist, D.R., "A Finite Element Analysis of Simple Panel Response to Turbulent Boundary Layers", AFFDL-TR-67-81, W-PAFB, OH, 1967.
26. Olson, M.D. and Lindberg, G.M., "Free Vibrations and Random Response of an Integrally-Stiffened Panel", National Research Council of Canada Report LR-544, Ottawa, Canada, 1970.
27. Olson, M.D., "A Consistent Finite Element Method for Random Response Problems", Computers and Structures, Vol. 2, 1972, pp. 163-180.
28. Penzien, J., Kaul, M.K., and Berge, B., "Stochastic Response of Off-Shore Towers to Random Sea Waves and Strong Motion Earthquakes", Computers and Structures, Vol. 2, 1972, pp. 733-756.
29. Hurty, W.C. and Rubinstein, M.F., "Dynamics of Structures", Prentice-Hall, 1964.
30. Rao, G.V., Raju, I.S., and Raju, K.K., "A Finite Element Formulation for Large Amplitude Flexural Vibrations of Thin Rectangular Plates", Computers and Structures, Vol. 6, 1976, pp. 163-167.
31. Yamaki, N., "Influence of Large Amplitudes on Flexural Vibrations of Elastic Plates", ZAMM, Vol. 41, 1961, pp. 501-510.
32. Wah, T., "Large Amplitude Flexural Vibration of Rectangular Plates", Int. J. Mech. Sci., Vol. 5, 1963, pp. 425-438.
33. Mei, C., "Nonlinear Vibration of Beams By Matrix Displacement Method", AIAA J., Vol. 10, March 1972, pp. 355-357.
34. Mei, C., "Finite Element Displacement Method for Large Amplitude Free Flexural Vibrations of Beams and Plates", Computers and Structures, Vol. 3, 1973, pp. 163-174.
35. Berger, H.M., "A New Approach to the Analysis of Large Deflections of Plates", JAM, Vol. 22, December 1955, pp. 465-472.
36. Rao, G.V., Raju, I.S. and Raju, K.K., "Nonlinear Vibrations of Beams Considering Shear Deformation and Rotary Inertia", AIAA J., Vol. 14, May 1976, pp. 685-687.
37. Raju, K.K. and Rao, G.V., "Nonlinear Vibrations of Orthotropic Plates by a Finite Element Method", J. Sound Vib., Vol. 48, 1976, pp. 301-303.

38. Rao, G.V., Raju, I.S. and Raju, K.K., "Finite Element Formulation for the Large Amplitude Free Vibrations of Beams and Orthotropic Circular Plates", Computers and Structures, Vol. 6, 1976, pp. 169-172.
39. Reddy, J.N. and Stricklin, J.D., "Large Deflection and Large Amplitude Free Vibrations of Thin Rectangular Plates using Mixed Isoparametric Elements", Symp. on Applications of Computer Methods in Engineering, University of Calif., Los Angeles, Calif, August 23-26, 1977.
40. Mei, C. and Rogers, J.R., Jr., "Application of the TRPLT1 Element to Large Amplitude Free Vibrations of Plates", NASA CP-2018, Oct. 1977, pp. 275-298.
41. Narayanaswami, R. and Mei, C., "Addition of Higher Order Plate Elements to NASTRAN", NASA TM X-3428, Oct. 1976, pp. 439-477.
42. Mei, C., Narayanaswami, R., and Rao, G.V., "Large Amplitude Free Flexural Vibrations of Thin Plates of Arbitrary Shape", Computers and Structures (submitted).
43. Cowper, G.R., Kosko, E., Lindberg, G.M. and Olson, M.D., "Static and Dynamic Applications of a High-Precision Triangular Bending Element", AIAA J., Vol. 7, 1969, pp. 1957-1965.
44. Raju, K.K. and Rao, G.V., "Calculation of Nonlinear Axisymmetric Vibrations of Thin Shells of Revolution by a Finite Element Method", J. Sound Vib., Vol. 38, 1975, pp. 505-509.
45. Raju, K.K. and Rao, G.V., "Large Amplitude Asymmetric Vibrations of Some Thin Shells of Revolution", J. Sound Vib., Vol. 44, 1976, pp. 327-333.
46. Evensen, D.A., "Comment on Large Amplitude Asymmetric Vibrations of Some Thin Shells of Revolution", J. Sound Vib., Vol. 52, 1977, pp. 453-455.

UNIFIED SYNTHESIS OF
MANUAL AND AUTOMATIC CONTROL
APPLIED TO INTEGRATED FIRE
AND FLIGHT CONTROL

by

David K. Schmidt

School of Aero And Astro
Purdue University

August 4, 1978

UNIFIED SYNTHESIS OF
MANUAL AND AUTOMATIC CONTROL
APPLIED TO INTEGRATED FIRE
AND FLIGHT CONTROL

David K. Schmidt*
Purdue University

Abstract

Given adequate open-loop specifications, for example, aircraft handling qualities criteria, design techniques are available to the system designer for synthesizing even the most complex flight control system. Unfortunately, however, weaknesses exist in the handling qualities specification, particularly for "non-conventional" vehicles such as control configured vehicles (CCV's). In this investigation, an augmentation method based on optimal control techniques and closed-loop, task-oriented criteria was applied to augment the system dynamics for a man-in-the-loop air-to-air tracking task. The cases of tracking during level flight (longitudinal tracking) as well as tracking during a high-banked maneuver was addressed. The predicted tracking performance with and without augmentation is compared to man-in-the-loop simulation results and the improved performance was significant.

* Assistant Professor of Aeronautics and Astronautics,
Purdue University

Table of Contents

I. Summary and Conclusions .	1
II. The Augmentation Methodology	3
III. Application to Longitudinal Tracking	9
IV. Analysis of Tracking During Maneuver	26
References	32
Appendices	
A. Line of Sight Equations	33
B. Lead Angle Equations	38
C. Target Kinematics	43
D. F-106 Perturbation Model - 4-g, Level Turn	47
E. Sample Output - Longitudinal Tracking	51

List of Figures

Figure 1, Eigenvalues - LCOS, 1000 ft.	14
Figure 2, Eigenvalues - LCOS, 3000 ft.	15
Figure 3, LCOS Augmentation Performance	18
Figure 4, Elevator Activity	19
Figure 5, Pilot's Stick Activity	19
Figure 6, Eigenvalues-Director, 1000 ft.	23
Figure 7, Eigenvalues-Director, 3000 ft.	24
Figure 8, Director Augmentation Performance	25

I. Summary and Conclusions

I.1 Summary

Given adequate open-loop specifications, for example, aircraft handling qualities criteria, design techniques, particularly modern control approaches, are available to the system designer for synthesizing even the most complex flight control systems. Unfortunately, however, weaknesses exist in the handling qualities areas, particularly for "non-conventional" aircraft such as control configured vehicles (CCV's). In this report, an augmentation synthesis method usable in the absence of quantitative handling qualities specifications will be presented, and the application of this method to augment the system dynamics for a man-in-the-loop air-to-air tracking task will be reported.

In the absence of open-loop specifications, the alternative is to design the augmentation via closed-loop performance criteria. With the use of an analytical pilot model, the augmentation synthesis and performance prediction will be obtained. However, in the type of problem being considered here in general, the selection of the type pilot model is important. Due to the predictive nature of models developed from optimal control theory and the applicability of optimal control theory to multi-dimensional, high-order system, this pilot modeling approach was adopted.

In the next section of the report, optimal-control pilot modeling is highlighted, and an augmentation method previously developed by the author (Ref. 1) is presented. The method involves simultaneously solving for the pilot model and augmentation. Simultaneous solution is important, and is required since the pilot is known to vary his equalization and gain subject to the plant dynamics, which is being determined with the augmentation process.

In Section III, the method is applied to a longitudinal tracking task, and several cases are investigated. Two ranges and two rms target acceleration levels are considered, with a lead-computing-optical sight (LOOS) as well as a perfect director sight. Initially, the predicted performance (for both sights) is compared with previously obtained experimental data, to establish pilot model credibility. Then a family of augmentation systems are developed for all cases cited above. With the full-state feedback systems so designed, significant tracking performance improvements are obtained. Additionally, the system dynamics are appreciably modified.

Finally, as an application of the pilot model to an even larger system, with multiple control inputs, the predicted piloted tracking performance for an F-106 vehicle in a level, four-g turning condition is compared with simulated results in Section IV. The simulation was recently performed in AFFDL's large amplitude motion simulator (LAMARS) facility. The results compared most favorably - a very important result considering the pilot model was a priori "tuned" with very nominal values rather than adjusted to fit the data.

A significant portion of this overall effort involved developing the linearized math models for the engagement geometry, line of sight, sight lead angles, target kinematics, and vehicle model at a large bank-angle condition. The results of this important effort are reported in Appendices A through D.

I.2 Conclusions

The favorable comparisons between predicted theoretical performance and simulated results was most gratifying, especially in the F-106 case. The intent of this latter case was to attempt to lend further credibility to the pilot modeling methods as a whole, and in the author's opinion this was accomplished. The data compared extremely well, as mentioned above, considering the lack of model adjustment necessary. Also, only one simulator run of approximately 30 seconds duration was obtained to compare with the steady state rms predictions. This is not therefore, representative of a large data base.

The augmentation results were most promising. The computer algorithm suggested work well without flaw in the solution of the coupled Riccati equations, and the performance improvements were significant. However, to proceed to a practical implementation of the methodology, the system performance with a full-state estimator, as well as with limited state feedback need to be evaluated. These two approaches are clearly practically realizable while the full-state, noise free approximations here developed constitute the first step in this synthesis approach.

II. The Augmentation Methodology

In the absence of quantitative dynamic specifications (i.e., handling qualities) for the vehicle in a specific task, the approach to be taken here will be to augment the vehicle such that an objective cost function based on the task is minimized. In addition, since the vehicle is manually controlled, the man-in-the-loop aspects must be explicitly included. Finally, the multi-dimensional, high-order system aspects must be considered. The approach will follow that of Ref. 1.

II.1 The Pilot Model

To include the man in the loop, an analytical model of the human controller is required, and the modeling approach to be followed here is to use the optimal control representation of the man rather than a describing function pilot model. The major drawback of a describing function model is that the pilot's gain and equalization is known to depend on the plant he is controlling (i.e., he is adaptive), and the plant dynamics are to be augmented, hence, are not known a priori. Furthermore, the optimal control model (OCM) is compatible with optimal control synthesis methods, well suited for multi-dimensional systems.

As presented in Reference 2, the optimal pilot model evolves from the assumption that the well-trained, well-motivated pilot selects his control input(s), u_p , subject to human limitations, such that the following objective is minimized,

$$J_p = E \left\{ \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (\bar{y}' Q \bar{y} + \bar{u}_p' R u_p + \dot{\bar{u}}_p' G \dot{\bar{u}}_p) dt \right\}$$

The dynamic system being controlled by the pilot is described by the familiar linear relation

$$\dot{\bar{x}} = A_p \bar{x} + B_p \bar{u}_p + \bar{w}$$

$$\bar{y} = C \bar{x} \tag{A}$$

where \bar{x} is the system state vector, \bar{u}_p the pilot control vector, \bar{y} the output vector, and \bar{w} is the vector of zero-mean external disturbances with covariance

$$E[\bar{w}(t)\bar{w}'(t+\sigma)] = W\delta(\sigma)$$

Included as human limitations are observation delay, τ , and observation noise, \bar{v}_y . So the pilot actually perceives the noise-contaminated,

delayed states, or

$$\bar{y}_p = C_p \bar{x}(t - \tau) + \bar{v}_y(t - \tau)$$

The covariances of the zero-mean observation noise may include the effects of perception thresholds and attention allocation, and is denoted

$$E[\bar{v}_y(t)\bar{v}_y'(t+\sigma)] = V_y \delta(\sigma)$$

Defining the augmented state vector, $\bar{x} = \text{col} [\bar{x}, \bar{u}_p]$, the solution to the problem, or the pilot's control is given as

$$\dot{\bar{u}}_p^* = -G^{-1}[0, I]K_p \bar{x}$$

where K_p is the positive definite solution to the Riccati equation

$$-\left[\begin{array}{c|c} A_p & B_p \\ \hline 0 & 0 \end{array} \right]' K_p - K_p \left[\begin{array}{c|c} A_p & B_p \\ \hline 0 & 0 \end{array} \right] - \left[\begin{array}{c|c} C_p' Q C_p & 0 \\ \hline 0 & R \end{array} \right] + K_p B_o G^{-1} B_o' K_p = \dot{K}_p$$

(B)

where $B_o' = [0, I]$

It will be convenient to partition K_p such that

$$K_p = \left[\begin{array}{c|c} K_{p1} & K_{p2} \\ \hline K_{p3} & K_{p4} \end{array} \right]$$

and note that now the equation for the optimal control \bar{u}_p^* is

$$\dot{\bar{u}}_p^* = -G^{-1}K_{p3} \hat{\bar{x}} - G^{-1}K_{p4} \bar{u}_p^*$$

or a linear feedback of the best estimate of the state, $\hat{\bar{x}}$, and some control dynamics. (These control dynamics have been shown to be equivalent to the pilot's neuro-muscular lag.)

To model the pilot's remnant, motor noise is usually added to the control equation. The final pilot's control is represented by

$$\dot{\bar{u}}_p^* = -G^{-1}K_{p3} \hat{\bar{x}} - G^{-1}K_{p4} \bar{u}_p^* + G^{-1}K_{p4} \bar{v}_m$$

where

$$E[\bar{v}_m(t)\bar{v}_m'(t+\sigma)] = V_M \delta(\sigma)$$

Now, the state estimator consists of a Kalman filter and a least-mean square predictor, or

$$\dot{\hat{x}}(t-\tau) = A_p \hat{x}(t-\tau) + \Sigma C_p' V_Y^{-1} [\bar{Y}_p(t) - C_p \hat{x}(t-\tau)] + B_p \bar{u}_p^*(t-\tau)$$

$$\hat{x}(t) = \bar{P}(t) + e \cdot A_p^T [x(t-\tau) - \bar{P}(t-\tau)]$$

with

$$\dot{\bar{P}} = A_p \bar{P} + B_p \bar{u}_p^*$$

and the estimation error covariance matrix Σ is the solution of the Riccati equation

$$A_p \Sigma + \Sigma A_p' + W - \Sigma C_p' V_Y^{-1} C_p \Sigma = [0]$$

This system of equations, when solved, determines the optimal-control pilot model. The actual numerical solutions were obtained with the program PIREP, documented in Reference 3.

II.2 Augmentation Synthesis Method

In the determination of the pilot model parameters above, we have expressed the system dynamics in terms of the matrices A_p and B_p . However, since the augmentation has not been defined, the augmented plant, A_p and B_p is as yet unknown.

Consider the un-augmented plant dynamics to be described by

$$\dot{\bar{x}} = A \bar{x} + B \bar{u} + \bar{w}$$

where, as before, \bar{x} is the system state vector and \bar{w} is the same disturbance vector. However, A and B are now the un-augmented system matrices, and \bar{u} is the control input vector. Now the total control input to the plant will include pilot input, \bar{u}_p , plus augmentation input, \bar{u}_{SAS} , or

$$\bar{u}_p = \bar{u} + \bar{u}_{SAS}$$

Further, from the pilot model, we know that although the feedback gains (e.g., $G_{p3}^{-1} K_{p3}$, $G_{p4}^{-1} K_{p4}$) have not been determined, the pilot's control input

is expressible as

$$\dot{\bar{u}} = -G_{p3}^{-1} K_{p3} \hat{\bar{x}} - G_{p4}^{-1} K_{p4} \bar{u}_p \quad (C)$$

Now, the estimate of the state, $\hat{\bar{x}}$, can be expressed in terms of the true state plus some estimation error, \bar{e} , or

$$\hat{\bar{x}} = \bar{x} + \bar{e}$$

By treating this error as another disturbance, \bar{w}_p , we can write the pilot's equation as

$$\dot{\bar{u}} = -G_{p3}^{-1} K_{p3} \bar{x} - G_{p4}^{-1} K_{p4} \bar{u}_p + \bar{w}_p$$

(Note, the disturbance term, \bar{w}_p , can also include the pilot's remnant as well.) Combining this relation with the plant dynamic and pilot equations we have

$$\dot{\bar{x}} = \begin{bmatrix} A & B \\ -G^{-1}K_{P3} & -G^{-1}K_{P4} \end{bmatrix} \bar{x} + \begin{bmatrix} B \\ 0 \end{bmatrix} u_{SAS} + \begin{bmatrix} \bar{w} \\ \bar{w}_p \end{bmatrix} \quad (D)$$

where

$$\bar{x} = \text{col}[\bar{x}, \bar{u}_p]$$

We may now proceed to determine an objective function for determining \bar{u}_{SAS} .

Since Hess (Ref. 4) has shown a strong correlation between pilot rating and the pilot's objective function we clearly see that the best (i.e., lowest) pilot rating implies the lowest pilot objective function. Therefore, for optimum pilot rating for a task, the control \bar{u}_{SAS} should be chosen to minimize J_p as defined in the pilot rating method. (This method defines the state^p and control weights, Q and R , as the inverse of the maximum allowable deviations in the variables as perceived by the pilot.) Finally, to preclude infinite augmentation gains, we must also penalize augmentation control energy. Therefore, the augmentation is chosen to minimize

$$J_{SAS} = J_p + E \left\{ \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \bar{u}_{SAS}' F \bar{u}_{SAS} dt \right\}$$

or

$$J_{SAS} = E \left\{ \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (\bar{y}' Q \bar{y} + \bar{u}_p' R \bar{u}_p + \dot{\bar{u}}_p' G \dot{\bar{u}}_p + \bar{u}_{SAS}' F \bar{u}_{SAS}) dt \right\}$$

and Q , R , and G are as chosen in the pilot's objective function, J_p . This may be written as

$$J_{SAS} = E \left\{ \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (\bar{x}' P \bar{x} + \bar{u}_{SAS}' F \bar{u}_{SAS}) dt \right\}$$

where

$$P = \begin{bmatrix} \bar{C}_p' Q C_p + K_{P3}' G^{-1} K_{P3} & K_{P3}' G^{-1} K_{P4} \\ K_{P4}' G^{-1} K_{P3} & R + K_{P4}' G^{-1} K_{P4} \end{bmatrix}$$

and instead of Equation 3 being substituted for $\dot{\bar{u}}_p$ in the above J_{SAS} , we have invoked a sort of separation principle and substituted the

relation

$$\dot{\bar{u}}_p = -G^{-1}K_{p_3}\bar{x} - G^{-1}K_{p_4}\bar{u}_p$$

The justification for this relation being used lies in the fact that we wish to synthesize the augmentation based on how the pilot is trying to perform the control function rather than on how the pilot is capable of doing so.

With this objective function and the system dynamics given in Equation D, the problem is now stated in conventional form, except K_{p_3} and K_{p_4} are as yet undetermined of course. If we assume, for example, full state feedback, the solution of this problem is known to be

$$\bar{u}_{SAS}^* = -F^{-1} \begin{bmatrix} B' & | & 0 \end{bmatrix} K_{SAS} \bar{x}$$

or

$$\bar{u}_{SAS}^* = -F^{-1}B'K_{SAS_1}\bar{x} - F^{-1}B'K_{SAS_2}\bar{u}_p$$

where

$$K_{SAS} = \begin{bmatrix} K_{SAS_1} & | & K_{SAS_2} \\ \hline K_{SAS_3} & | & K_{SAS_4} \end{bmatrix}$$

is the solution to the Riccati equation

$$-\begin{bmatrix} A & | & B \\ \hline -G^{-1}K_{p_3} & | & -G^{-1}K_{p_4} \end{bmatrix}, \quad K_{SAS} - K_{SAS} \begin{bmatrix} A & | & B \\ \hline -G^{-1}K_{p_3} & | & -G^{-1}K_{p_4} \end{bmatrix}$$

$$-P + K_{SAS} \begin{bmatrix} B \\ \hline 0 \end{bmatrix} F^{-1} \begin{bmatrix} B' & | & 0 \end{bmatrix} K_{SAS} = \dot{K}_{SAS}$$

(E)

We see in this expression that the solution for K_{SAS} obviously depends on K_p (or K_{p_3} and K_{p_4}). Returning to the Riccati equation for

the pilot gain (Equation B), we also see that that equation depends in turn on the SAS gains (or K_{SAS}) since the pilot Riccati equation involves the augmented plant matrices A_p and B_p . As a result of the SAS design procedure just presented, we now know, however, the SAS structure. Returning to the pilot model, we may now include this SAS structure specifically, so that A_p and B_p (as in Equation A) may in fact be expressed as

$$A_p = A - BF^{-1}B'K_{SAS_1}$$

and

$$B_p = B(I - F^{-1}B'K_{SAS_2}) \quad (F)$$

Substituting these expressions in the pilot Riccati equation yields two coupled Riccati equations, one for the pilot gains, Equation B, and one for the SAS gains, Equation E. These may be solved simultaneously for K_{SAS} and K_p . The equations to be integrated are obtained by multiplying out the submatrices in Equations B and E. Then the elements of the submatrices K_{p_1} , K_{SAS_1} , etc. are then found by integrating the resulting 8 matrix equations backwards in time.

Now the pilot model is used to evaluate augmented system performance with the plant matrices (A_p and B_p) found from Eqn. F.

III Application to Longitudinal Tracking- LCOS and Director

The initial application of the methodology involved the augmentation of the tracking task in the longitudinal axis only. This piloting task was first investigated with the optimal pilot model by Harvey and Dillow (Ref. 7). In the cited investigation the dynamic model included the longitudinal, short-period approximation for four different aircraft and a lead-computing optical sight (LCOS). The task was experimentally investigated in the laboratory using an analog simulator equipped with oscilloscope (display) and force-sensitive stick control. The optimal pilot model was then applied to "predict" the manual control performance. (Actually, in this case the experimental data was used to tune some of the pilot model parameters and compare predicted performance.)

III.1 PIREP VALIDATION

To assure the validity of the PIREP program a set of cases considered by Harvey were evaluated with the PIREP code. The math model development for this case is as follows.

The F-4E aircraft dynamics at an altitude of 15,000 ft. and a MACH number of 0.9 ($U_0 = 952$ fps) were used. The stability derivatives in this case are

$$\begin{aligned} Z_{\dot{\alpha}} &= -1.033 U_0 \text{ (ft/sec}^2\text{)} & M_{\dot{\alpha}} &= -.344 \text{ (rad/sec)} \\ Z_{\delta} &= -.0951 U_0 \text{ (ft/sec}^2\text{)} & M_{\delta} &= -.738 \text{ (rad/sec)} \\ M_{\dot{\alpha}} &= -10.44 \text{ (rad/sec}^2\text{)} & M_{\delta} &= -37.1 \text{ (rad/sec}^2\text{)} \end{aligned}$$

In addition, the elevator dynamics were modeled in this case with the relation

$$\dot{\delta}_E = -20\delta_E + 16\delta_{stick}$$

With the above parameters, the longitudinal equations for the short period approximation are for this case

$$\begin{aligned} \dot{\theta} &= q \\ \dot{\alpha} &= Z_{\alpha}\alpha + q + Z_{\delta}\delta_E \\ \dot{q} &= (M_{\alpha} + M_{\dot{\alpha}}Z_{\alpha})\alpha + (M_q + M_{\dot{\alpha}})q \\ &\quad + (M_{\delta} + M_{\dot{\alpha}}Z_{\delta})\delta_E \\ \dot{\delta}_E &= -20\delta_E + 16\delta_{stick} \end{aligned}$$

(A)

The trim condition is, of course, level flight ($\phi_1 = \dot{\phi}_1 = P_1 = Q_1 = R_1 = V_1 = W_1 = 0$),

$\alpha_1 = 10^\circ$ (Ref. 8).

As can be verified from Equations H in Appendix B, the sight equation for the perturbation elevation lead angle, λ_{EL} , is

$$\dot{\lambda}_{EL} = -\frac{1}{T_f} \lambda_{EL} - \left(\frac{Z\alpha}{2V} + \frac{J_v V_A}{D} \right) \alpha + \dot{\theta} \quad B$$

where the steady state lead angle and line of sight are assumed zero, and the attacker's normal acceleration, a_z , is approximately equal to $Z\alpha$. Furthermore, in this case the attacker's and target's steady-state velocities (e.g., U_1, V_1) and specific accelerations (A_{A_z}, A_{T_z} ; etc.) are assumed equal. The ballistic coefficient, J_v , was estimated from a relation given in Harvey (Ref. 7) as

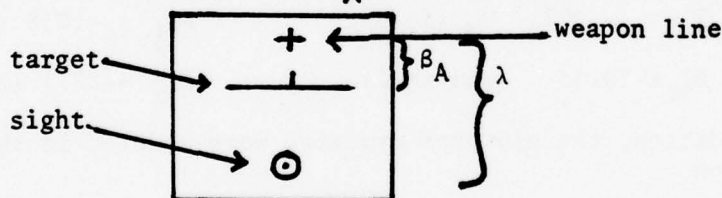
$$J_v = .0382$$

(Note that the relation in Harvey Ref. 7 does not agree with the definition of J_v given in Ref. 5.) Finally, the time of flight, T_f , was estimated with the method given in Ref. using a muzzle velocity, V_m , of 3400 ft/sec. Since two ranges ($D = 1000$ ft and 3000 ft) were considered, two times of flight were estimated. They were $T_f = .327$ sec ($V_f = 2110$ ft/sec) for $D = 1000$ ft, and $T_f = 1.30$ sec ($V_f = 1350$ ft/sec) for $D = 3000$ ft.

The relation used by Harvey for the pipper error, ϵ , is

$$\epsilon = \lambda - \beta_A$$

as shown in the sketch, where β_A is the relative line-of-sight angle.



Display Sketch

It's dynamics were described by the relation

$$\beta_A = \theta - \beta_I$$

where β_I is the inertial line of sight rate, and

$$\dot{\beta}_I = \frac{V_T}{D} \gamma_T + \frac{V_A}{D} (\alpha - \theta)$$

$$\gamma_T = -\frac{1}{V_T} a_{z_T}$$

In the above relation, γ_T is the target flight path angle and a_{z_T} is the target's normal acceleration described by the Markov process

$$\dot{n}_1 = -\frac{1}{\tau_a} n_1 + \xi$$

$$\dot{a}_T = -\frac{1}{\tau_a} a_T + n_1$$

The target time constant chosen was 3 seconds and σ_{a_T} was selected as 3.5 and 5.0 g's. Since (Ref.)

$$\sigma_{a_T}^2 = \frac{\tau_a^3 \sigma_\xi^2}{4},$$

selection of σ_{a_T} specifies the noise statistics. The math model for four different conditions is now complete; two distances, 1000 and 3000 ft, and two target acceleration levels, 3.5 and 5.0 g's rms.

The pilot model parameters used are given in Table 1 and the rms performance comparisons are given in Table 2. As a result of the close agreement, PIREP was considered satisfactory. A sample output is given in Appendix E.

Table 1
Pilot Model Parameters

Observation delay, $\tau = 0.2$ sec.

Neuromuscular time constant, $\tau_N = .1$ sec.

Observation vector, $\bar{Y}' = [\epsilon, \dot{\epsilon}, \lambda, \dot{\lambda}]$

Cost function weights, $Q_{y_{ii}} = [16., 1., 0., 4.]$

$Q_u = 0.$

Observation noise, $V_{y_i} = \pi \rho \sigma_y^2$; $\rho = .01$

Motor noise, $V_u = \pi \rho \sigma_u^2$; $\rho = .0015$ (Harvey)
 $\rho = .001$ (Schmidt)

Observation thresholds, $Th_\epsilon = Th_\lambda = 0.65^\circ$
 $Th_\epsilon = Th_\lambda = 0.65^\circ$

Full Attentional Allocation

Table 2
Performance Comparison - LCOSS

	RMS Parameter			
	Pipper Error, ϵ (deg)	Lead Angle, λ (deg)	Pitch Rate, q (deg/sec)	Elevator $\delta\epsilon$ (deg)
1000 ft $\sigma_T = 3.5$ g -----				
Experiment	2.1	2.9	8.0	2.1
Harvey	2.3	2.9	8.8	2.1
Schmidt	2.3	2.9	8.7	2.1
1000 ft $\sigma_T = 5$ g -----				
Experiment	2.7	4.0	11.2	3.0
Harvey	3.0	4.2	12.4	3.0
Schmidt	3.0	4.2	12.1	3.0
3000 ft., $\sigma_T = 3.5$ g -----				
Experiment	2.4	8.7	6.4	1.8
Harvey	2.4	9.2	7.6	1.9
Schmidt	2.3	10.1	7.3	1.8
3000 ft., $\sigma_T = 5$ g -----				
Experiment	2.7	12.1	9.2	2.4
Harvey	2.9	13.0	10.5	2.6
Schmidt	2.7	14.2	10.0	2.5

III.2 Augmentation of Tracking with LCOS

As shown in Section II, the augmentation, assuming full-state feedback, is determined from the relation

$$\begin{aligned} \bar{u}_{SAS} &= -F'B'K_{SAS_1}\bar{x} - F'B'K_{SAS_2}\bar{u}_p \\ \text{or} \quad \bar{u}_{SAS} &= -K_x\bar{x} - K_\delta\delta_{stick} \end{aligned} \quad C.$$

where the system dynamics are described by

$$\dot{\bar{x}} = A\bar{x} + B\bar{u} + w$$

In the longitudinal tracking case previously discussed, the state vector may be chosen as

$$\bar{x}' = [n_1, a_{T_2}, \gamma_T, \beta_T, \lambda, \alpha, \theta, q, \delta_E]$$

and the pilot's control is $u_p = \delta_{stick}$. After selection of the SAS weighting, F , the SAS and pilot's Riccati equations are solved simultaneously, as noted in Section II, to determine K_{SAS_1} and K_{SAS_2} .

The gains K_x and K_δ in Eqn. C are, of course, now available.

For the LCOS tracking cases presented in the previous section, a family of gains (K_x and K_δ) are given in Table 3. Note that since the two tracking distances are considered, we have two sets of system dynamics, hence two sets of gains.

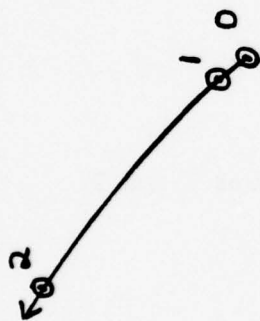
The augmented plant matrices are

$$A_p = A - BK_x$$

$$B_p = B(1 - K_\delta)$$

The eigenvalues of the plant matrix A_p are shown in Figures 1 and 2.

(3) $-12.64 \pm 11.1j$ - 3



System Eigen values - LCOS, 1000 FT:

X = Open-Loop Roots

\odot = Piloted Closed-Loop Roots

O - Un-Augmented

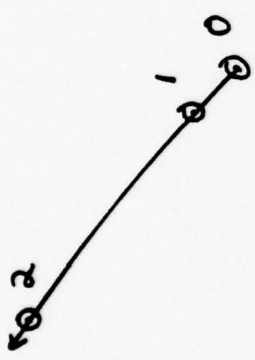
$$1 - F_{SAS}^{-1} = 10^{-3}$$

$$2 - F_{SAS}^{-1} = 10^{-2}$$

$$3 - F_{-1} = 10^{-1}$$



Figure 1 - Eigenvalues - LCOS , 1000 Ft



(3) $-12.5 \pm 11.2j$

System Eigenvalues - LCOS, 3000 Ft.

X - Open Loop Roots

o - Piloted Closed-Loop Roots

O - Un-Augmented

1 - $F_{SAS}^{-1} = 10^{-3}$

2 - $F_{SAS}^{-1} = 10^{-2}$

3 - $F_{SAS}^{-1} = 10^{-1}$

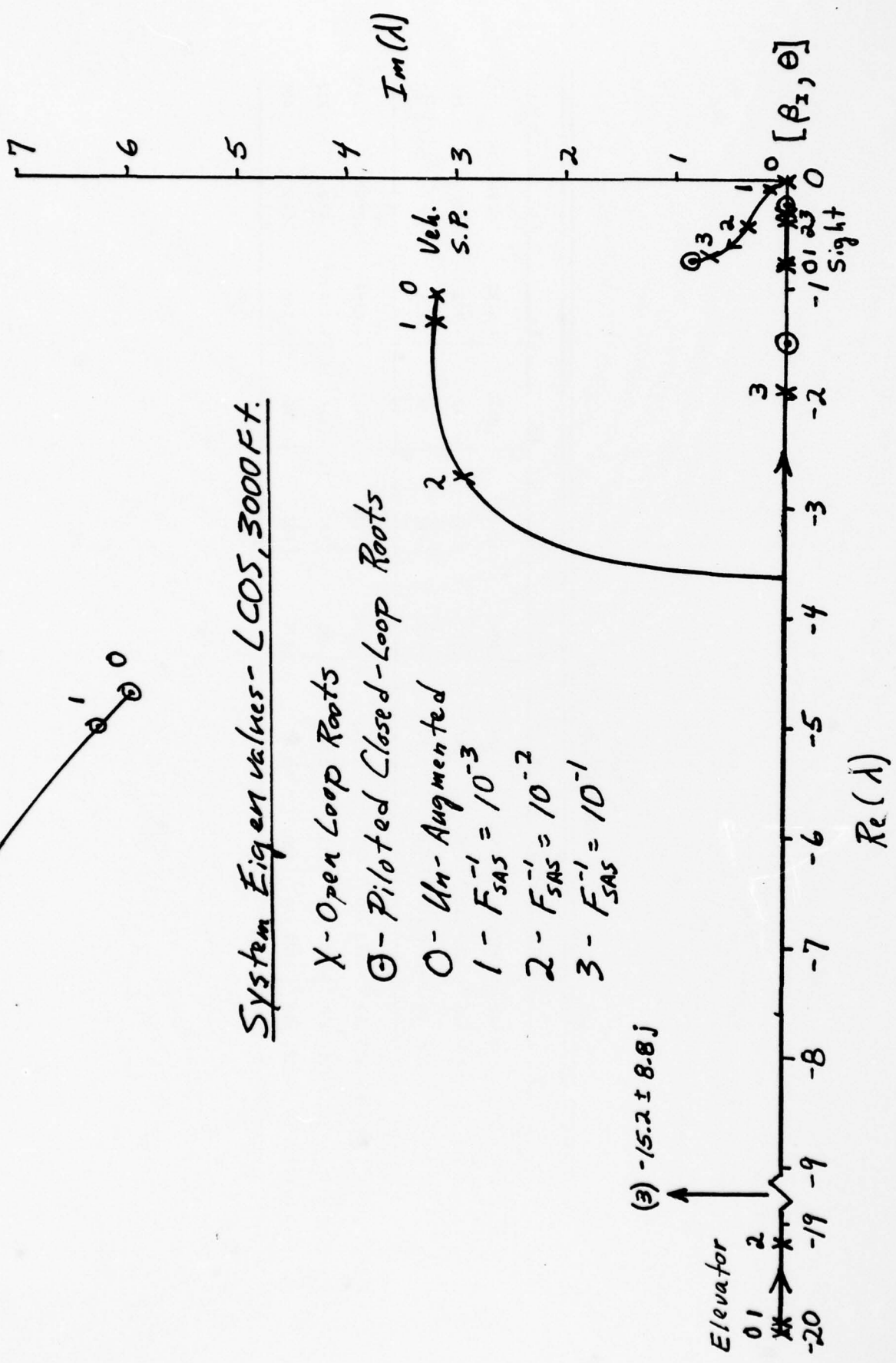


Figure 2, Eigenvalues - LCOS, 3000 Ft

The predicted performance (pipper error and elevator activity) of the piloted system is depicted in Figure 3 as a function of augmentation level for both a 3.5g and 5g rms target acceleration. Also shown for reference is the unaugmented system performance ($F^{-1} = 0$). Clearly, with increased augmentation level, tracking accuracy is reduced while total elevator activity remains almost constant.

This last result was somewhat surprising until an analysis of the various components of commanded elevator was performed. Recall that the SAS control input is

$$u_{SAS} = -K_x \bar{x} - K_{\delta} \delta_{stick\ pilot}$$

where u_{SAS} , analogous to the pilot's control, is an equivalent stick input. Therefore, the commanded elevator deflection, δ_{E_c} , is

$$\delta_{E_c} = K_g \delta_{stick}$$

where K_g is a gearing ratio (.8 in this case). Now we write

$$\begin{aligned} \delta_{E_{cSAS}} &= -K_g (K_x \bar{x} + K_{\delta} \delta_{st.p}) \\ \text{or} \\ \delta_{E_{cSAS}} &= -K_g K_x \bar{x} - K_g K_{\delta} \delta_{st.p} \end{aligned}$$

Since our results showed positive values for K_{δ} , we see that the SAS is cancelling some of the pilot's commanded elevator.

A plot of the total elevator deflection, pilot effective commanded elevator deflection $(1-K_{\delta}) K_g \delta_{st.p}$ and effective SAS elevator deflection $-K_g K_x \bar{x}$ is shown in Figure 4. And the total pilot's stick activity, $\delta_{st.p}^g$, is shown in Figure 5 for the 1000 ft., 5g case.

As can be seen, the augmentation input is simply replacing the pilot's input. And since in this analysis full-state feedback and no augmentation sensor errors are assumed, the SAS will always be more precise than the pilot. A first estimate of the optimal gains would be the case for $F^{-1} = .01$.

III.3 Tracking and Augmentation with Perfect Director

As discussed in Appendix B as well as Ref. 5, a perfect director sight is based on the assumption of perfect knowledge of the target's acceleration and the line-of-sight rates, rather than using the attacker's parameters to estimate these values as in LCOS. The effect is a sight with much less dynamics. To investigate the effect of a different sight, the analysis was also performed with a director.

Effect of Augmentation on Performance

----- Pitch Error
----- Elevator Defl.

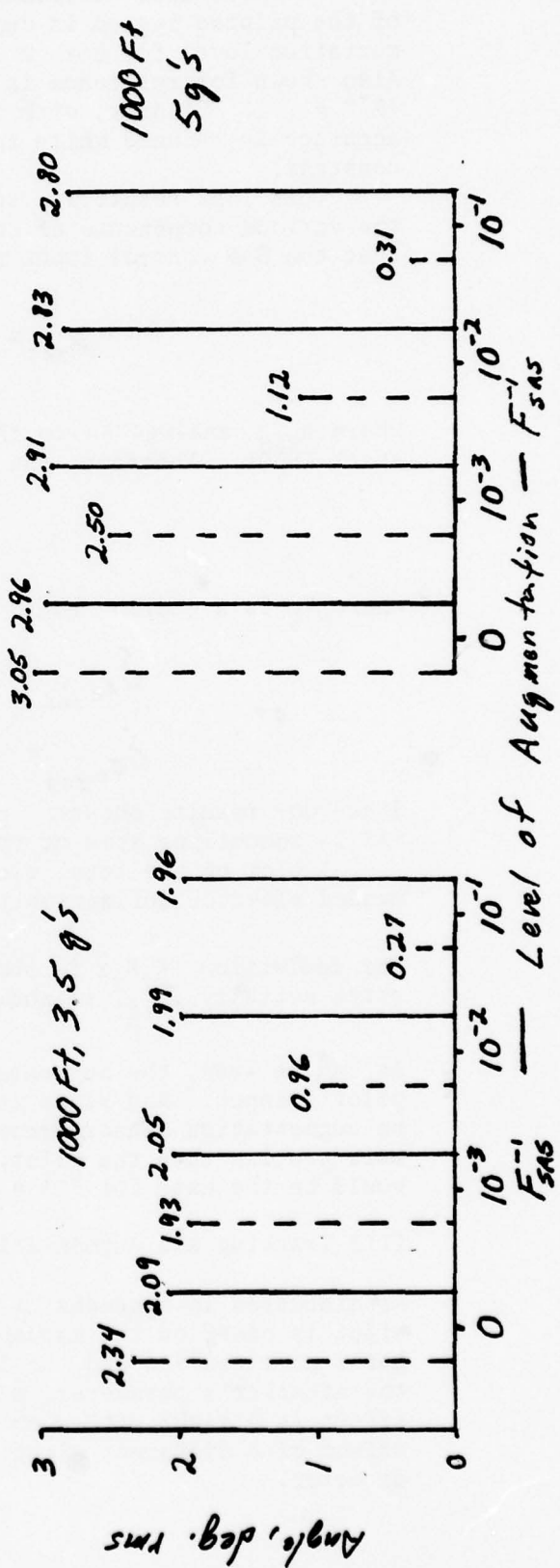
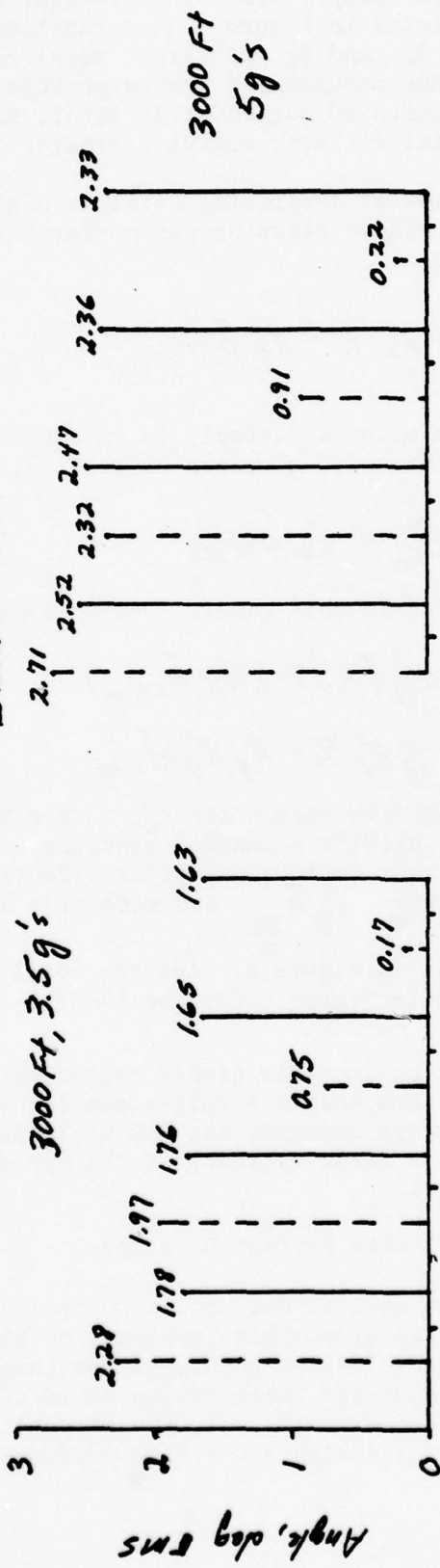


Figure 3, LCOS Augmentation Performance

Elevator Activity Breakdown

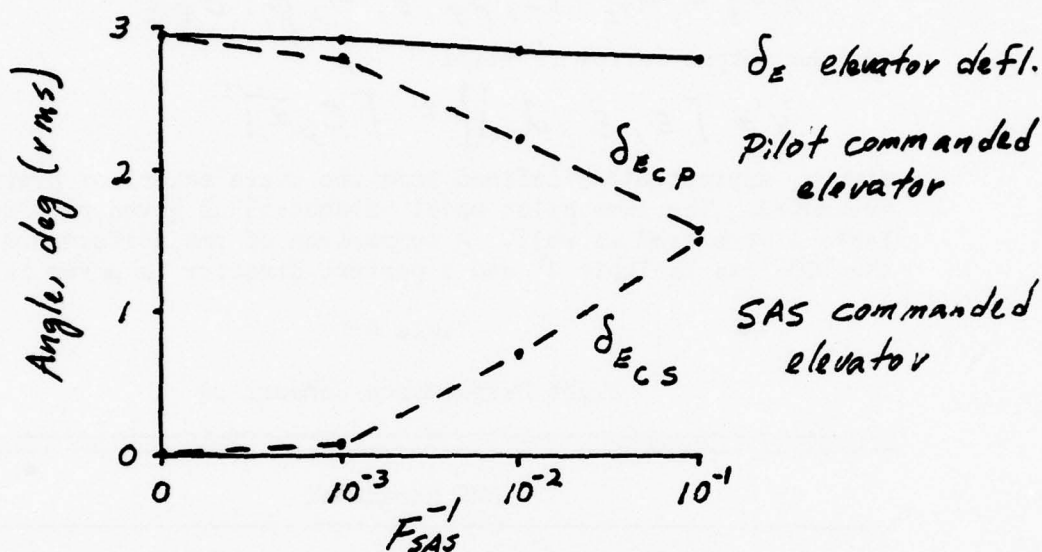
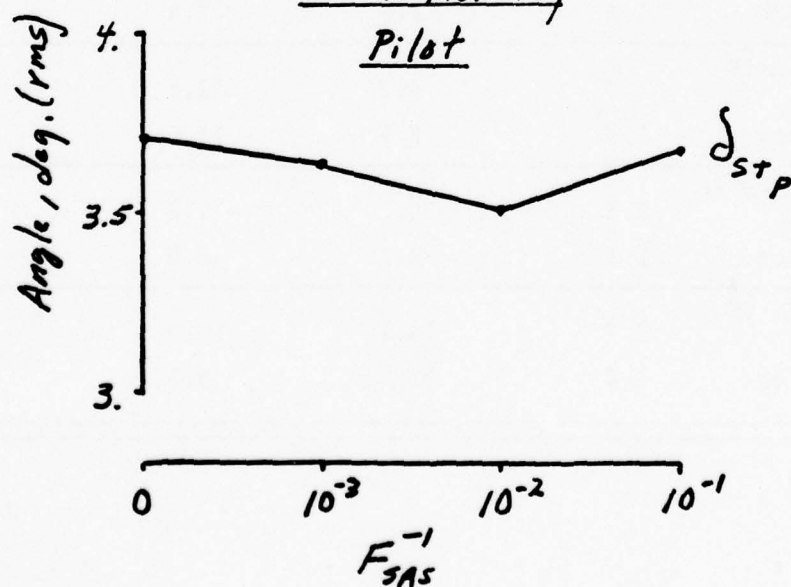


Figure 4 - Elevator Activity

Stick Activity Pilot



Level of Augmentation

Figure 5 - Pilot's Stick Activity

In this case, the equation for the lead angle (Eqn. B) is replaced by the relation

$$\lambda = T_f \dot{\beta}_I - \frac{T_f}{2V_f} a_{Tz} - J_v \frac{V_A}{V_f} \alpha$$

As a result, the state vector becomes

$$\bar{x}' = [n, a_{Tz}, \gamma_T, \beta_I, \alpha, \theta, \phi, \delta_E]$$

and the output vector is still

$$\bar{y}' = [\epsilon, \dot{\epsilon}, \lambda, \dot{\lambda}] = [C_p \bar{x}]'$$

with c_p appropriately defined from the state equations previously presented. The same pilot model parameters as given previously in Table 1 were used as well. A comparison of rms performance between the LCOS (as in Table 2) and a perfect director is given in Table 4.

Table 4

Sight Performance Comparison

Case	RMS Parameter			
	Pipper Error, ϵ (deg)	Lead Angle, λ (deg)	Pitch Rate, q (deg/sec)	Elevator Defl. (deg)
<u>1000ft., 3.5g</u>				
LCOSS	2.3	2.9	8.7	2.1
Director	1.6	2.6	7.9	2.0
<u>1000ft., 5g</u>				
LCOSS	3.0	4.2	12.1	3.0
Director	1.9	3.7	11.0	2.9
<u>3000ft., 3.5g</u>				
LCOS	2.3	10.1	7.3	1.8
Director	1.1	9.7	6.0	1.6
<u>3000ft., 5g</u>				
LCOS	2.7	14.2	10.0	2.5
Director	1.4	13.8	8.5	2.3

The pilot-optimal augmentation analysis was performed for the director cases as well, and the optimal gains are given in Table 5, where, as in the LCOS case,

$$u_{SAS} = -K_x \bar{x} - K_\delta \delta_{stick}$$

with

$$K_x = F^{-1} B' K_{SAS1},$$

$$K_\delta = F^{-1} B' K_{SAS2}$$

Further, the augmented plant matrices become

$$A_p = A - B K_x$$

$$B_p = B(1 - K_\delta)$$

and the plant eigenvalues are shown in Figures 6 and 7.

Comparing these figures with Figures 1 and 2 indicates large differences in the effect of augmentation on the vehicle short period mode. This important difference highlights the effect of total system dynamics (pilot, sight, etc.) on the desired vehicle dynamics. In the case for $F^{-1} = .01$, for example, with the LCOS a significant change in short-period damping with little change in natural frequency. However, with the director, the natural frequency is measurably increased. Finally, changing the LCOS sight time constant (T_s) significantly altered the eigenvalue locus.

As can be seen in Figure 8, finally, the tracking performance is significantly improved, although not as dramatically as in LCOS (Figure 3), and the elevator trend is as before.

Table 5

Optimal Gains - Perf. Director

	K_{n1}	K_{aT}	K_{dT}	K_{β}	K_{α}	K_{θ}	K_q	K_{δ}	K_{δ_s}
$\frac{1000 \text{ ft}}{F^{-1}} = .001$	$-.113 \times 10^{-5}$	$-.735 \times 10^{-5}$.0209	.0301	.0295	-.0510	-.0105	.0168	.0286
$F^{-1} = .01$	$-.725 \times 10^{-5}$	$-.492 \times 10^{-4}$.144	.213	.189	-.357	-.072	.111	.154
$F^{-1} = .1$	$-.267 \times 10^{-4}$	$-.206 \times 10^{-3}$.632	.995	.734	-1.63	-.314	.432	.383
$\frac{3000 \text{ ft}}{F^{-1}} = .001$	$-.611 \times 10^{-5}$	$-.185 \times 10^{-4}$.0164	.0302	.0249	-.0467	-.0106	.0171	.0289
$F^{-1} = .01$	$-.406 \times 10^{-4}$	$-.129 \times 10^{-3}$.115	.214	.159	-.329	-.073	.112	.155
$F^{-1} = .1$	$-.168 \times 10^{-3}$	$-.589 \times 10^{-3}$.522	.997	.610	-1.52	-.316	.435	.384

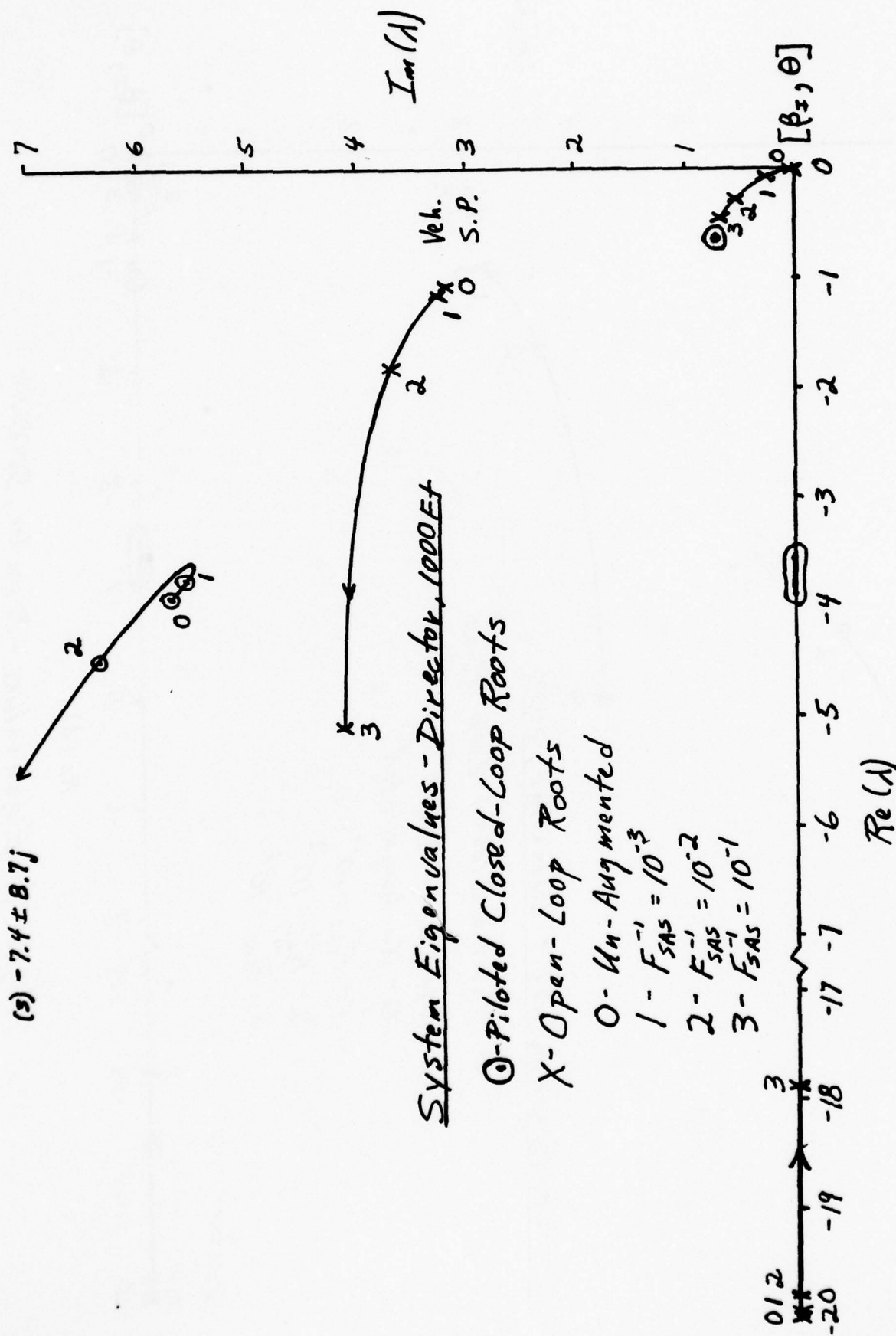
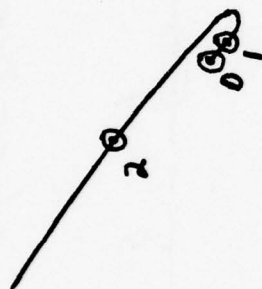


Figure 6. Eigenvalues - Director, 1000 Ft

$$(3) -7.4 \pm 8.7j$$



System Eigenvalues - Director, 3000

○ - Piloted Closed-Loop Roots

X - Open-Loop Roots

○ - Un-Augmented

$$1 - F_{SAS}'' = 10^{-1}$$

$$2 - F_{SAS}' = 10^{-2}$$

$$3 - F_{SAS}' = 10^{-1}$$

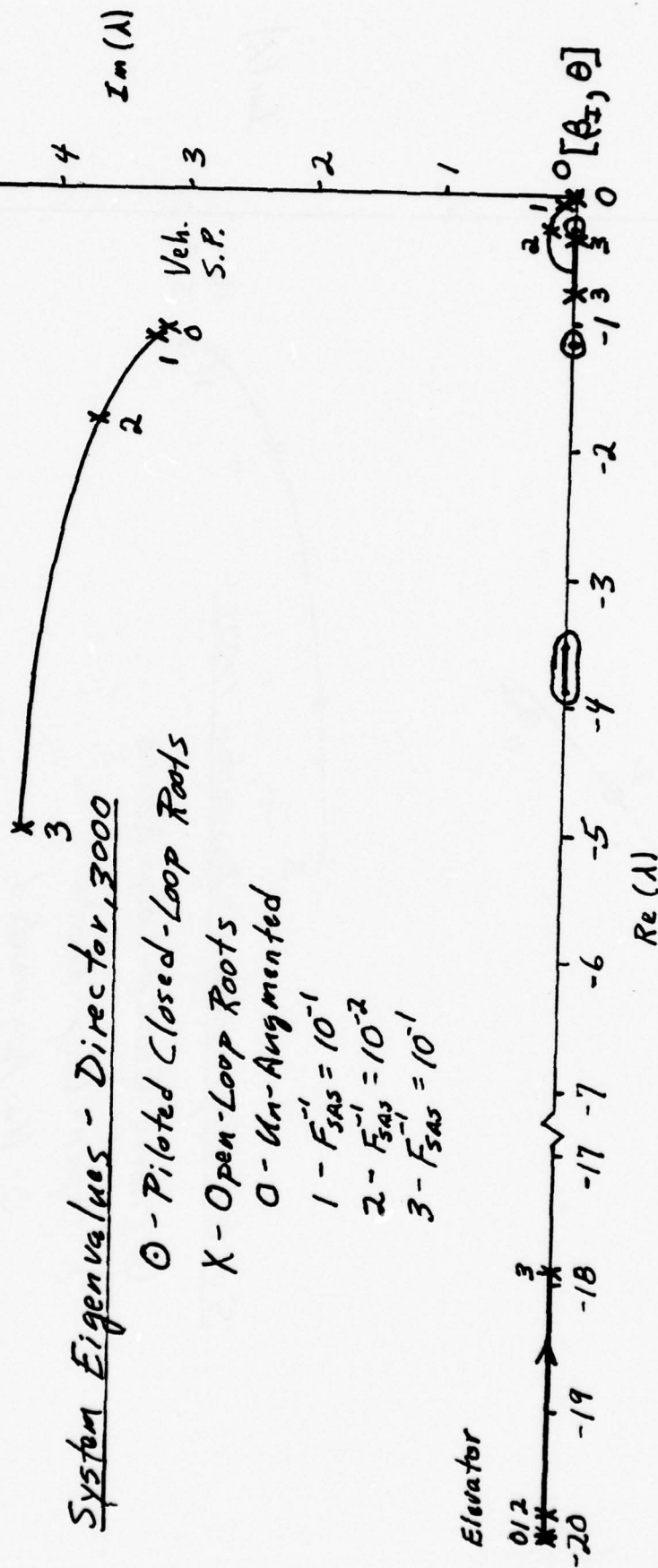


Figure 7, Eigenvalues - Director, 3000 Ft

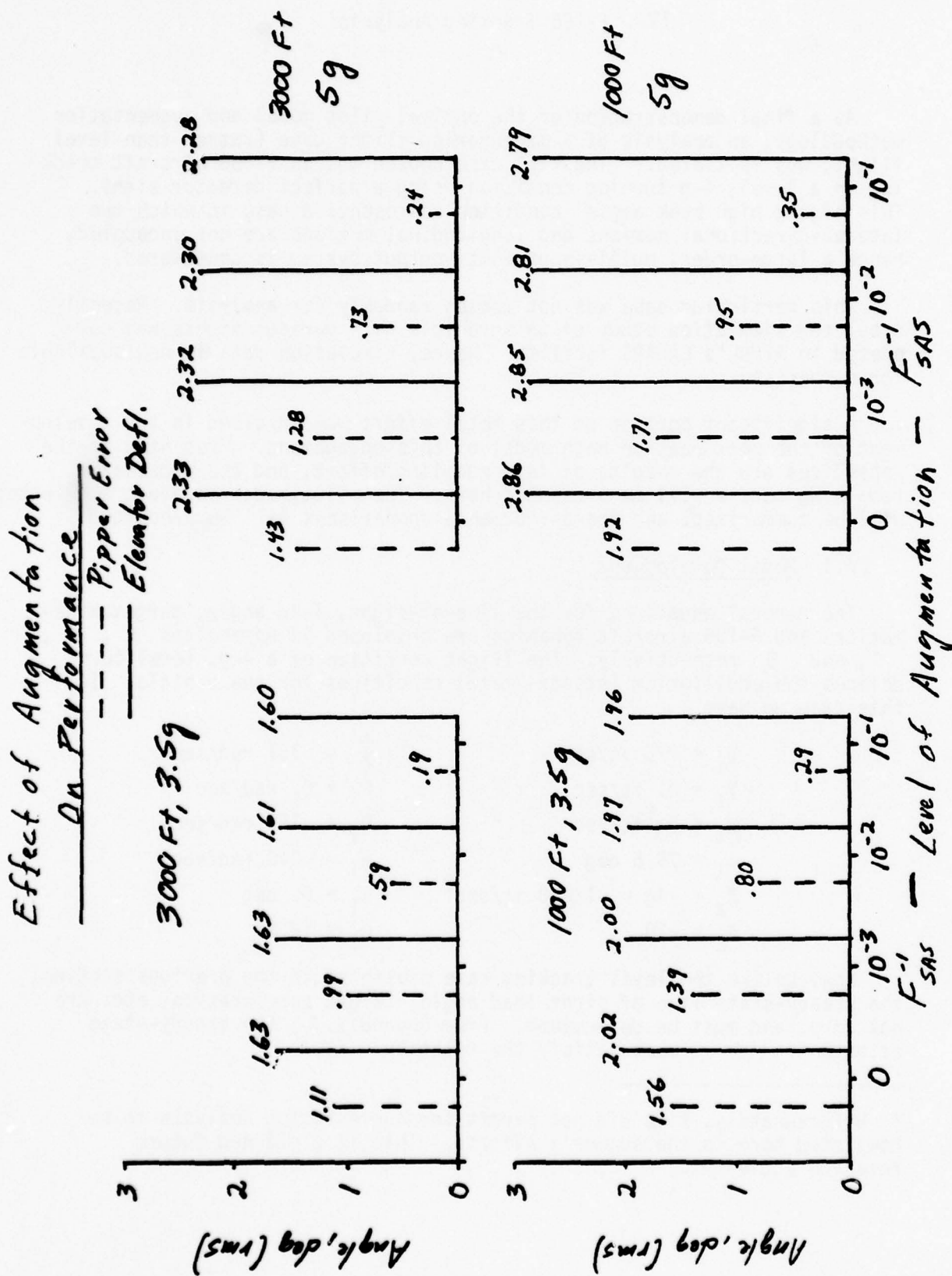


Figure 8, Director Augmentation Performance

IV. F-106 Tracking Analysis

As a final demonstration of the optimal pilot model and augmentation methodology, an analysis of a maneuvering flight case (rather than level flight) was initiated.* The test case chosen was an F-106 aircraft tracking in a level, 4-g turning condition using a perfect director sight. This high-g high bank angle condition represents a case in which the lateral-directional motions and longitudinal motions are not uncoupled, hence a large-order, multi-input multi-output system is considered.

This particular case was not chosen randomly for analysis. Recently, a piloted simulation study of this vehicle with various sights was completed in AFFDL's LAMARS facility. Hence, simulation data is now available for comparison.

A significant portion of this total effort was involved in the development of the perturbation math model of this engagement. Presented in the Appendices are the results of this modeling effort, and the necessary remaining models will be presented here. The pilot model parameters selected will be summarized, and the performance comparisons will be presented.

IV.1 Model Development

The general equations for the line-of-sight, lead angle, target kinematics, and F-106 aircraft dynamics are developed in Appendices A, B, C, and D respectively. The flight condition of a 4-g, level turn defines the equilibrium (steady-state) conditions for the vehicle. In this case we have

$U_1 = 775 \text{ ft/sec}$	$\dot{\psi} = .161 \text{ rad/sec}$
$V_1 = 0. \text{ ft/sec}$	$P_1 = 0. \text{ rad/sec}$
$W_1 = 0. \text{ ft/sec}$	$Q_1 = .156 \text{ rad/sec}$
$\phi_1 = 75.5 \text{ deg}$	$R_1 = .040 \text{ rad/sec}$
$A_z = -4g = -128.8 \text{ ft/sec}^2$	$\beta_1 = 0. \text{ deg}$
$\delta_E = -20.2^\circ$	$\alpha_1 = 14.5^\circ$

Now, unlike the level tracking case presented in the previous section, the steady-state line of sight, lead angle, target acceleration, etc. are not zero, and must be determined. From Appendix A, the steady-state azimuth LOS must satisfy the relation

* Unfortunately, time did not permit the augmentation analysis to be completed here in the summer's efforts. This is a planned future research activity.

$$D_1 \dot{B}_{Az} = 0 = (V_{T_1} - V_T) - B_{Az} [\cos \alpha_1 (U_{T_1} - U_1) - \sin \alpha_1 (W_{T_1} - W_1)] \\ - D_1 [R_1 (\cos \alpha_1 - B_{E1} \sin \alpha_1) + P_1 (\sin \alpha_1 + B_{E1} \cos \alpha_1)] \quad (A)$$

Now, from geometry, we can write the expressions for the target velocity in the attacker's stability coordinate system as

$$U_{T_1} = V_T \cos \Delta\psi$$

$$V_{T_1} = V_T \sin \Delta\psi \cos \phi_A$$

$$W_{T_1} = -V_T \sin \Delta\psi \sin \phi_A$$

where

$$V_T = \text{target's velocity}$$

$$\Delta\psi = \text{target-attacker heading difference } (\psi_T - \psi_A)$$

$$\phi_A = \text{attacker's bank angle}$$

Substituting into equation A and noting that in steady state $V_A = V_T$ and $P_1 = 0$ yields

$$\frac{R_1 D_1}{V_A} (\cos \alpha_1 + \sin \alpha_1 B_{EL}) = -B_{Az} \cos \alpha_1 (\cos \Delta\psi - 1.) \\ + \sin \Delta\psi (\cos \phi_1 - B_{Az} \sin \alpha_1 \sin \phi_1) \quad (B)$$

Now, assume a nominal tracking range D_1 of 2000 ft, $\sin \alpha_1 B_{EL} \ll \cos \alpha_1$, and $B_{Az} \approx 0$ yields

$$\sin \Delta\psi_1 \approx .4 \\ \Delta\psi_1 \approx 23.6^\circ$$

Therefore, the target's velocity and accelerations are (see Appendix C)

$$\begin{array}{ll} U_{T_1} = 710 \text{ ft/sec} & A_{T_{x_1}} = -50. \text{ ft/sec}^2 \\ V_{T_1} = 78 \text{ ft/sec} & A_{T_{y_1}} = 0. \text{ ft/sec}^2 \\ W_{T_1} = -320 \text{ ft/sec} & A_{T_{z_1}} = -118. \text{ ft/sec}^2 \end{array} \quad (C)$$

Where the target is assumed to be in a level, 4-g turn also.

Now, the bullet time of flight, lead angles, and line of sight may be estimated. From Ref. 5, the projectile flight distance, time of flight, average velocity and attacker velocity are related by

$$T_f (V_A + V_f) = D_f$$

and

$$D_f \approx D + (V_A + \dot{D}) T_f$$

where

D is the present range.

Therefore, the average flight velocity is

$$V_f \approx D/T_f$$

Finally, using Ref. 5, if $V_{\text{muzzle}} = 3300$ ft/sec, the time of flight is estimated as $T_f \approx .786$ sec, $V_f \approx 2545$ ft/sec, and $J_v = .185$. From Appendix B, the steady lead-angle equations for the director are then

$$L_{EL}(2545) = B_{EL}(5.2) + 323.$$

If, for zero pipper error, $B_{EL} \approx L_{EL}$, we have $L_{EL} \approx B_{EL} = .127$ rad.

Likewise for azimuth,

$$L_{AZ}(-2545) = -B_{AZ}(5.2) + 78$$

If $B_{AZ} \approx L_{AZ}$, we have $L_{AZ} \approx -.031$ rad. Re-iterating, assuming zero steady-state pipper error, we have

$$L_{EL} = B_{EL} = .127 \text{ rad}$$

$$L_{AZ} = B_{AZ} = -.031 \text{ rad}$$

(D)

With these steady values, and those given in Equation C, we now have specified the coefficients of the perturbation equations for range and line-of-sight rate (Appendix A), lead angles (Appendix B), attacker kinematics (Appendix C), and vehicle dynamics (Appendix D).

Now, choose the seventeen-element state vector as

$$\bar{x}' = [n_1, a_T, n_2, \phi_T, \Delta\psi, v_T, \omega_T, d, \beta_{E1}, \beta_{AZ}, \theta, \phi, \alpha, q, \beta, p, r]$$

and the equations just cited can be used to form the linear model

$$\dot{\bar{x}} = A \bar{x} + B \bar{U}_p + E \bar{w}$$

with $\bar{U}_p' = [\delta_{E_{\text{stick}}}, \delta_{A_{\text{stick}}}]$

and target noises

$$\bar{w}' = [\xi_1, 0, \xi_2, 0, \dots]$$

(Note that the rudder pedal has not been included as a control. Discussions with fighter pilots indicate little rudder usage in this tracking task. This addition could be made later however. Also, note that the attacker and target perturbations U and U_T are omitted to limit the order of the state vector.)

With this model, and steady-state conditions as defined, the system eigenvalues are as follows:

Short period, $- 0.91 \pm 3.33j$
 Dutch roll, $- 0.43 \pm 4.87j$
 Roll, $- 1.0$
 Spiral, $- .002$
 Target accel., $- 1.0 \pm 0.j$
 Target roll, $-1.0 \pm 0.j$
 Lead angles, $\left\{ \begin{array}{l} -.005 \pm .16j \\ \text{line of sight} \\ \text{and kinematics} \end{array} \right. \begin{array}{l} -.006 \pm .16j \\ 3 \text{ remaining} = 0. \end{array}$

IV.2 Performance Comparison

Two different simulation runs were performed in a level turn with a perfect director sight. Since the steady-state, or mean values of the performance parameters differed between the two runs, as shown in Table 6, it was decided to disregard run number 109 results. The justification for this is that the mean values from this run did not agree with the stated flight condition (e.g., for a level, 4-g turn, trim bank-angle ϕ_1 is 75.5°).

The parameters selected for this pilot model are shown in Table 7. It should be noted that these parameters are nominal values used frequently in other pilot modeling investigations (see Hess, Ref. B), except that the output weighting matrix, Q_y , is the same as that used previously in the longitudinal tracking analysis.

Table 6
Steady-State Parameter Comparison

	Lead Angles λ_{EL} λ_{Az} (rad.)		Target Velocities (gun coord.) v_T (fps) w_T (fps)		
Run 109	.025	-.028	27.	-65.	
Run 133	.123	-.048	100.	-312.	
Theoretical	.127	-.031	78.	-302.	
	Euler Angles (body) ϕ_1 θ_1 (deg)		Body Rates P_1 Q_1 R_1 (deg/sec)		
Run 109	28.9	3.4	38.0	2.9	2.6
Run 133	72.6	3.6	- .6	8.3	2.5
Theoretical	75.5	3.6	- .6	8.9	2.2
	Velocity Angles α_1 β_1 (deg)		Stick Deflections $\delta_{E_{st}}$ $\delta_{A_{st}}$ (in)		
Run 109	3.8	0.6	-1.	-1.	
Run 133	10.9	0.6	-2.3	- .4	
Theoretical	14.5	0.	-1.0	-0.	

The observation (output) vector includes pipper error and lead angles, as before, plus the target bank angle. This angle is considered an important pilot cue, but because of the range, its observation threshold is increased.

Table 7
Pilot Model Parameters

Observation delay, $\tau = .2$ sec
Neuromuscular lag, $\tau_{N_E} = \tau_{N_A} = .2$ sec
Motor noise, $V_{u_i} = \pi \rho_m \sigma_{u_i}^2$, $\rho_E = \rho_A = .01$ (-20 dB)
Observation noise, $V_{y_i} = \frac{\pi \rho_i}{f_i} \sigma_{y_i}^2$, $\rho_i = .01$ (-20dB)
Fractional attention, 0.5 to longitudinal and lat-dir axis, $f_i = .5$, all i
Output Vector, $\bar{y}' = [\epsilon_{E1}, \dot{\epsilon}_{E1}, \epsilon_{Az}, \dot{\epsilon}_{Az}, \lambda_{E1}, \dot{\lambda}_{E1}, \lambda_{Az}, \dot{\lambda}_{Az}, \phi_T, \dot{\phi}_T]$
Observation thresholds, .05 deg, .1 deg/sec except for ϕ_T - 5 deg, 10 deg/sec
Output weights, Q_y ; $Q_\epsilon = 16.$, $Q_{\dot{\epsilon}} = 1.$, $Q_\lambda = 4.$

The angle off ($\approx \Delta\psi$) was not included because in this in-plane encounter, the pilot was not considered to perceive this variable accurately.

Before comparison of the rms performances, it is important to point out one fundamental difference between the simulation and the model prediction method. The model is structured by assuming the pilot is performing a regulator task, while the target is randomly maneuvering. On the other hand, this was not the situation simulated. The task simulated was a capture and track, that is, it was more a variable initial condition task. Furthermore, the simulated target exhibited essentially no rms accelerations and bank angle about the mean. Consequently, the model target accelerations (in this case, bank angle) must simply be guessed, then parametric comparisons made with these facts in mind. One must compare relative to v_T and ω_T which can vary greatly with the time constants selected in the target acceleration and bank angle noise model. And these time constants must be selected purely on judgment. Clearly, after comparison, target time constants could be adjusted, but the intent here was to be "predictive" as possible. The performance is given in Table 8. It's felt that the agreement is more than satisfactory.

Table 8
RMS Performance Comparison

	Target Motion		Target Velocities*		
	σ_{a_t} (fps)	σ_{ϕ_T} (deg)	N_T	W_T (ft/sec)	
Simulated	0.	0.3	34.	48.	
$\tau_{\phi_T} = 5^\circ$	0.	5.0	28.	121.	
$\tau_{\phi_T} = 7^\circ$	0.	7.0	38.	164.	
	Tracking Errors		Lead Angles		
	ϵ_{Az}	ϵ_{E1} (rad)	λ_{Az}	λ_{E1} (rad)	
Simulated	.017	.019	.045	.030	
$\tau_{\phi_T} = 5^\circ$.007	.024	.012	.048	
$\tau_{\phi_T} = 7^\circ$.009	.032	.016	.064	
	Euler Angles (body)		Body Rates		
	ϕ	θ (deg)	p	q	r (deg/sec)
Simulated	4.2	1.1	5.0	1.7	0.9
$\tau_{\phi_T} = 5^\circ$	3.5	4.4	1.8	0.4	0.9
$\tau_{\phi_T} = 7^\circ$	4.9	6.1	2.5	0.6	1.2
	Velocity Angles		Stick Deflections		
	α	β (deg)	$\delta_{E_{st}}$	$\delta_{A_{st}}$ (in)	
Simulated	1.4	0.2	0.3	0.2	
$\tau_{\phi_T} = 5^\circ$	0.3	0.1	0.1	0.1	
$\tau_{\phi_T} = 7^\circ$	0.4	0.2	0.1	0.1	

* Depend on σ_{ϕ_T} and time constants.

References

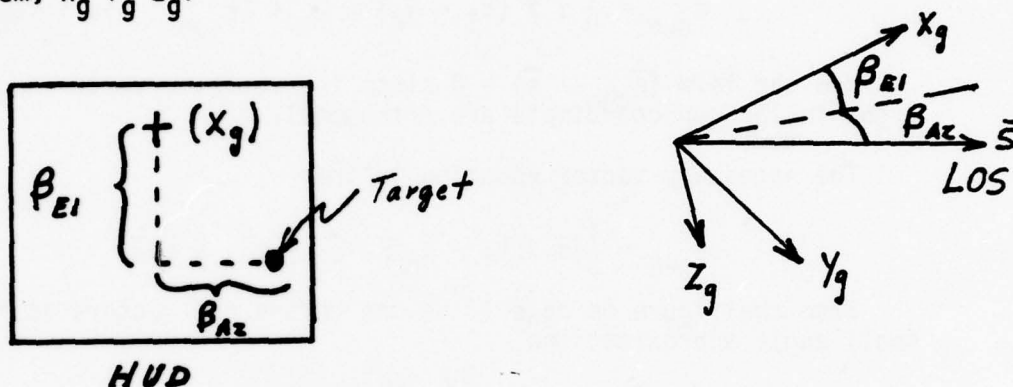
1. Schmidt, D.K., "Optimal Flight Control Synthesis Via Pilot Modeling," AIAA Paper No. 78-1286, AIAA Guid. and Control Conference, Palo Alto, California, Aug. 1978.
2. Kleinman, D.L., S. Baron, and W.H. Levison, "An Optimal Control Model of Human Response, Part I: Theory and Validation," and Part II: Prediction of Human Performance in a Complex Task," Automatica, Vol.6, 1970, pp. 357-383.
3. Curry, R.E., et. al., "Pilot Modeling for Manual Simulation," AFFDL-TR-76-124, Vol.I & II. December, 1976.
4. Hess, R.A., A Method for Generating Numerical Pilot Opinion Ratings Using the Optimal Pilot Model, NASA-TMX-73, 101, Feb., 1976.
5. Hohwiesner, W., Capt., "Principles of Airborne Fire Control," USAF Academy Notes, Dept. of Astro. and Computer Sciences, USAFA, Dec., 1975.
6. Hess, R.A., "Application of a Model-Based Flight Director Design Technique to a Longitudinal Hover Task," Journal of Aircraft, Vol. 14, No. 3, March, 1977.
7. Harvey, T.R., "Application of an Optimal Control Pilot Model to Air to Air Combat," AFIT M.S. Thesis, GA/MA/74M-1, March, 1974
8. Roshkam, J. Flight Dynamics of Rigid and Elastic Airplanes, Vol. 1, Lawrence, Kansas, 1972.
9. "Elasticized Stability And Control Aerodynamic Coefficients For The F-106A and F-106B Airplanes ...", Convair Report No. DA -SB-219, 31 October 1958.

Appendix A

Line of Sight Equations

A.1 Vector Equation Development

Clearly, the angle(s) to the line of sight (LOS) are required for our analysis, and as usual, linearized perturbation equations are ultimately required. The equations to be developed in this section will be derived for the angles, as shown in the figure below, in the vehicle's gun coordinate system, $X_g Y_g Z_g$.



Now the kinematics of the line of sight unit vector, \bar{s} , are described by

$$\bar{V}_{\text{Target}} - \bar{V}_{\text{Attacker}} = \frac{d}{dt}(D\bar{s})$$

where D is the scalar distance from the attacker to the target. Therefore,

$$\bar{V}_T - \bar{V}_A = \frac{d}{dt}(D\bar{s}) = \dot{D}\bar{s} + D\dot{\bar{s}}$$

A.

But the time rate of change of a unit vector is the cross-product of that vector with its rate of rotation vector in inertial space, $\dot{\bar{\beta}}$, or

$$\dot{\bar{s}} = \dot{\bar{\beta}} \times \bar{s}$$

So $\dot{\bar{\beta}}$ is the rate of rotation of the LOS in the inertial (e.g. earth-fixed) coordinate system.

Taking the cross-product of Eqn. A with \bar{s} yields

$$\bar{s} \times (\bar{V}_T - \bar{V}_A) = \bar{s} \times \dot{D}\bar{s} + \bar{s} \times (\dot{\bar{\beta}} \times \bar{s}) D$$

And from the vector triple product identity

$$\bar{s} \times (\dot{\bar{\beta}} \times \bar{s}) = \dot{\bar{\beta}} - (\dot{\bar{\beta}} \cdot \bar{s}) \bar{s}$$

Now, instead of the inertial LOS rate $\dot{\bar{\beta}}$, we desire the LOS rate in the attacker's gun coordinate system or $\dot{\bar{\beta}}_{gun}$, which is

$$\dot{\bar{\beta}}_{gun} = \dot{\bar{\beta}} - \bar{\omega}_g$$

where $\bar{\omega}_g$ is the rotation rate vector of the gun coordinate system. Substituting into the above yields

$$\dot{\bar{\beta}}_{gun} + \bar{\omega}_g - [(\dot{\bar{\beta}}_{gun} + \bar{\omega}_g) \cdot \bar{s}] \bar{s} = \frac{1}{D} \bar{s} \times (\bar{V}_T - \bar{V}_A)$$

or

$$\dot{\bar{\beta}}_{gun} = \frac{1}{D} \bar{s} \times (\bar{V}_T - \bar{V}_A) - \bar{\omega}_g + (\dot{\bar{\beta}}_{gun} \cdot \bar{s}) \bar{s} + (\bar{\omega}_g \cdot \bar{s}) \bar{s}$$

Now the term $(\dot{\bar{\beta}}_{gun} \cdot \bar{s}) \bar{s} = 0$ since the rotation vector and the line of sight in the gun coordinate are orthogonal.

The necessary vector equation is then

$$\dot{\bar{\beta}}_{gun} = \frac{1}{D} [\bar{s} \times (\bar{V}_T - \bar{V}_A)] - \bar{\omega}_g + (\bar{\omega}_g \cdot \bar{s}) \bar{s} \quad \text{B.}$$

From the figure on page 1, we can define the vectors as follows, using small angle approximations

$$\bar{s} = \bar{i}_g + \beta_{Az} \bar{j}_g + \beta_{El} \bar{k}_g$$

$$\bar{\beta}_g = -\beta_{El} \bar{j}_g + \beta_{Az} \bar{k}_g$$

$$\bar{\omega}_g = P_g \bar{i}_g + Q_g \bar{j}_g + R_g \bar{k}_g$$

$$\bar{V}_A = U_A \bar{i}_g + V_A \bar{j}_g + W_A \bar{k}_g$$

$$\bar{V}_T = U_T \bar{i}_g + V_T \bar{j}_g + W_T \bar{k}_g$$

(It's important to note that \bar{V}_T and the components here defined are the velocity components of the target in the attacker's gun coordinate system.) Carrying out the operations in Equation B., and assuming $\bar{\omega}_g \cdot \bar{s} \approx P_g$, we have the desired relations

$$\dot{\beta}_{El} = \frac{1}{D} [(W_T - W_g) - \beta_{El} (U_T - U_g)] + Q_g - P_g \beta_{Az} \quad \text{C.}$$

$$\dot{\beta}_{Az} = \frac{1}{D} [(V_T - V_g) - \beta_{Az} (U_T - U_g)] - R_g + P_g \beta_{El}$$

To obtain the equation for the rate of change of range, or \dot{D} , we can simply take the dot product of Equation A. with \bar{s} to obtain

$$(\bar{V}_T - \bar{V}_A) \cdot \bar{s} = \dot{D} (\bar{s} \cdot \bar{s}) + D (\dot{\bar{\beta}} \times \bar{s}) \cdot \bar{s}$$

so

$$\dot{D} = (\bar{V}_T - \bar{V}_A) \cdot \bar{s}$$

$$\dot{D} = (U_T - U_g) + (V_T - V_g) \beta_{Az} + (W_T - W_g) \beta_{El} \quad \text{D.}$$

A.2 Scalar Equations in Stability Axes

Since all our equations are ultimately expressed in the vehicle stability axes, we need to express the above equations in terms of stability-axes variables rather than gun-coordinate variables.

Assuming the gun axis to be aligned with the vehicle body reference axis, we have the following relations between gun $(\cdot)_g$ and stability $(\cdot)_s$ variables

$$U_g = U_s \cos \alpha_1 - W_s \sin \alpha_1$$

$$V_g = V_s$$

$$W_g = W_s \cos \alpha_1 + U_s \sin \alpha_1$$

$$P_g = P_s \cos \alpha_1 - R_s \sin \alpha_1$$

$$Q_g = Q_s$$

$$R_g = R_s \cos \alpha_1 + P_s \sin \alpha_1$$

The resulting equations are

$$\begin{aligned} \dot{D} = & [(U_{Ts} - U_s) \cos \alpha_1 - (W_{Ts} - W_s) \sin \alpha_1] \\ & + \beta_{Az}(V_{Ts} - V_s) \\ & + \beta_{El}[(W_{Ts} - W_s) \cos \alpha_1 + (U_{Ts} - U_s) \sin \alpha_1] \end{aligned}$$

$$\begin{aligned} \dot{\beta}_{El} = & \frac{1}{D}[(W_{Ts} - W_s) \cos \alpha_1 + (U_{Ts} - U_s) \sin \alpha_1] \\ & - \beta_{El}\{(U_{Ts} - U_s) \cos \alpha_1 - (W_{Ts} - W_s) \sin \alpha_1\} \\ & + Q_s - \beta_{Az}(P_s \cos \alpha_1 - R_s \sin \alpha_1) \end{aligned}$$

E.

$$\begin{aligned} \dot{\beta}_{Az} = & \frac{1}{D}[(V_{Ts} - V_s) - \beta_{Az}\{(U_{Ts} - U_s) \cos \alpha_1 - (W_{Ts} - W_s) \sin \alpha_1\}] \\ & - (R_s \cos \alpha_1 + P_s \sin \alpha_1) + \beta_{El}(P_s \cos \alpha_1 - R_s \sin \alpha_1) \end{aligned}$$

where U_{Ts} , V_{Ts} , W_{Ts} are now the components of the target velocity in the attacker's stability axis and α_1 is the attacker's trim angle of attack (or the trim angle between the gun axis and stability axis).

A.3 The Perturbation Equation

Now to linearize the above equations, we will define the motion variables

in terms of a constant, steady-state quantity $(\cdot)_1$ and a time-varying perturbation about the steady-state $(\cdot)_1$ variable. So, let

$$\begin{aligned} D &= D_1 + d & U_S &= U_1 + u & P_S &= P_1 + p \\ \beta_{EL} &= \beta_{EL1} + \beta'_{EL} & V_S &= V_1 + v & Q_S &= Q_1 + q \\ \beta_{AZ} &= \beta_{AZ1} + \beta'_{AZ} & W_S &= W_1 + w & R_S &= R_1 + r \end{aligned}$$

Inserting these into Equations E will yield two sets of equations, one for the steady state values and one for the perturbation quantities. For example, taking the β_{E1} equation we have

$$\begin{aligned} (D_1 + d)(\dot{\beta}_{E1_1} + \dot{\beta}'_{E1}) &= [(W_{T_1} - W_1) \cos \alpha_1 + (U_{T_1} - U_1) \sin \alpha_1] \\ &\quad + [(W_T - w) \cos \alpha_1 + (U_T - u) \sin \alpha_1] \\ &\quad - (\beta_{E1_1} + \beta'_{E1})[(U_{T_1} - U_1) \cos \alpha_1 - (W_{T_1} - W_1) \sin \alpha_1] \\ &\quad + (U_T - u) \cos \alpha_1 - (W_T - w) \sin \alpha_1] \\ &\quad + (D_1 + d)(Q_1 + q) \\ &\quad - (D_1 + d)(\beta_{AZ_1} + \beta'_{AZ})[(P_1 + p) \cos \alpha_1 - (R_1 + r) \sin \alpha_1] \end{aligned} \quad F.$$

Now the steady state $(\cdot)_1$ values certainly must obey the original equation, so

by definition

$$\begin{aligned} D_1 \dot{\beta}_{E1_1} &= [(W_{T_1} - W_1) \cos \alpha_1 + (U_{T_1} - U_1) \sin \alpha_1] \\ &\quad - \beta_{E1_1} [(U_{T_1} - U_1) \cos \alpha_1 - (W_{T_1} - W_1) \sin \alpha_1] \\ &\quad + D_1 Q_1 - D_1 \beta_{AZ_1} [P_1 \cos \alpha_1 - R_1 \sin \alpha_1] \end{aligned} \quad G.$$

We may, as a result subtract this relation out of the complete, original relation (Eqn. F). In addition, under the assumption that the perturbation variables are small, we will drop higher order terms (e.g., products of pert. quantities). We are left with

$$\begin{aligned} D_1 \dot{\beta}'_{E1} + d \dot{\beta}_{E1_1} &= [(W_T - w) \cos \alpha_1 + (U_T - u) \sin \alpha_1] \\ &\quad - \beta_{E1_1} [(U_T - u) \cos \alpha_1 - (W_T - w) \sin \alpha_1] \\ &\quad - \beta'_{E1} [(U_{T_1} - U_1) \cos \alpha_1 - (W_{T_1} - W_1) \sin \alpha_1] \\ &\quad + D_1 q + Q_1 d - d \beta_{AZ_1} (P_1 \cos \alpha_1 - R_1 \sin \alpha_1) \\ &\quad - \beta'_{AZ} D_1 (P_1 \cos \alpha_1 - R_1 \sin \alpha_1) - D_1 \beta_{AZ_1} (p \cos \alpha_1 - r \sin \alpha_1) \end{aligned} \quad H.$$

This equation is now linear in terms of the perturbation variables, and the steady-state variables are constant by definition.

In like manner, the remaining two equations in E result in

$$\ddot{\beta}_1^0 = [(U_{T1} - U_1) \cos \alpha_1 - (W_{T1} - W_1) \sin \alpha_1] + B_{AZ}(V_{T1} - V_1) + B_{E1} [(W_{T1} - W_1) \cos \alpha_1 + (U_{T1} - U_1) \sin \alpha_1]$$

$$\dot{d} = [(u_T - u) \cos \alpha_1 - (w_T - w) \sin \alpha_1] + \beta_{AZ}'(V_{T1} - V_1) + B_{AZ1}(v_T - v) + \beta_{E1}'[(W_{T1} - W_1) \cos \alpha_1 + (U_{T1} - U_1) \sin \alpha_1] + B_{E11}[(w_T - w) \cos \alpha_1 + (u_T - u) \sin \alpha_1]$$

$$\ddot{\beta}_{AZ1}^0 = \frac{1}{D_1}[(V_{T1} - V_1) - B_{AZ}\{(U_{T1} - U_1) \cos \alpha_1 - (W_{T1} - W_1) \sin \alpha_1\}] (R_1 \cos \alpha_1 + P_1 \sin \alpha_1) + B_{E11}(P_1 \cos \alpha_1 - R_1 \sin \alpha_1)$$

and

$$\begin{aligned} D_1 \ddot{\beta}_{AZ} = & (v_T - v) - B_{AZ1}[(u_T - u) \cos \alpha_1 - (w_T - w) \sin \alpha_1] \\ & - \beta_{AZ}'[(U_{T1} - U_1) \cos \alpha_1 - (W_{T1} - W_1) \sin \alpha_1] \\ & - D_1(r \cos \alpha_1 + p \sin \alpha_1) + d(R_1 \cos \alpha_1 + P_1 \sin \alpha_1) \\ & + D_1 B_{E11}(p \cos \alpha_1 - r \sin \alpha_1) + D_1 \beta_{E1}'(P_1 \cos \alpha_1 - R_1 \sin \alpha_1) \\ & + d B_{E11}(P_1 \cos \alpha_1 - R_1 \sin \alpha_1) \end{aligned}$$

B. Lead Angle Equations

B.1 Perfect Director

Again linearized equations are needed for the lead angles computed by the sights. From Ref. 5, page 6-9, the fundamental equation for the lead angle, $\bar{\lambda}$, in general, is

$$V_f \bar{\lambda} = D[\dot{\bar{\beta}} - (\dot{\bar{\beta}} \cdot \bar{s})\bar{s}] + J_V V_A \bar{\alpha}_{\text{gun}} + (\bar{s} \times \dot{\bar{V}}_T) \frac{T_f}{2} \quad (A)$$

where, as defined in the above reference

$$\bar{\lambda} = \bar{s} \times \bar{i}_{\text{gun}} = \lambda_{E1} \bar{j}_{\text{gun}} - \lambda_{Az} \bar{k}_{\text{gun}}$$

V_f = average projectile velocity in still air

$[V_f = D_f/T_f - V_A, D_f = \text{distance of flight}]$

T_f = projectile time of flight

V_A = attacker's velocity

J_V = Jump correction factor, $= \frac{V_M - V_f}{V_A + V_M}$, V_M is projectile muzzle velocity

$\bar{\alpha}_{\text{gun}} = \bar{i}_{\text{gun}} \times \frac{\bar{V}_A}{|\bar{V}_A|} = -\alpha \bar{j}_{\text{gun}} + \beta \bar{k}_{\text{gun}}$, α and β here are angle of attack and sideslip

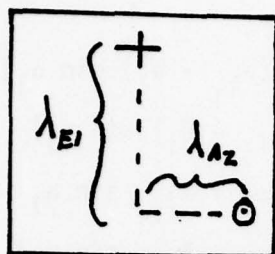
$\dot{\bar{V}}_T$ = attacker's acceleration

Now, if the actual line-of-sight rates are assumed known as in a director sight, from the LOS equations given previously

$$D[\dot{\bar{\beta}} - (\dot{\bar{\beta}} \cdot \bar{s})\bar{s}] = \bar{s} \times (\bar{V}_T - \bar{V}_A)$$

Furthermore, if the targets acceleration, $\dot{\bar{V}}_T$, is assumed known, we have the fundamental equation for a perfect director in terms of the target's velocity and acceleration.

To obtain the scalar component equations, define the following vectors in the gun coordinate system



HUD

$$\bar{\lambda} = \lambda_{E1} \bar{J}_{gun} - \lambda_{Az} \bar{k}_{gun}$$

$$\bar{\alpha}_{gun} = -\alpha \bar{J}_{gun} + \beta \bar{k}_{gun}$$

$$\dot{\bar{V}}_T = A_{T_x} \bar{i}_{gun} + A_{T_y} \bar{j}_{gun} + A_{T_z} \bar{k}_{gun}$$

$$\bar{V}_T = U_{T_g} \bar{i}_{gun} + V_{T_g} \bar{j}_{gun} + W_{T_g} \bar{k}_{gun}$$

where as in the LOS equation development, it's important to remember that the target's velocity and acceleration components are in the attacker's gun coordinate system. Carrying out the vector operation yields

$$V_f \lambda_{E1} = [\beta_{E1}(U_{T_g} - U_g) - (W_{T_g} - W_g)] - J_V V_A \alpha + (\beta_{E1} A_{T_x} - A_{T_z}) \frac{T_f}{2} \quad (B)$$

$$-V_f \lambda_{Az} = [(V_{T_g} - V_g) - \beta_{Az}(U_{T_g} - U_g)] + J_V V_A \beta + (A_{T_y} - \beta_{Az} A_{T_y}) \frac{T_f}{2}$$

Now, these equations will be expressed in terms of stability axis velocities, etc. rather than gun axis. We have the relations

$$U_g = U_s \cos \alpha_1 - W_s \sin \alpha_1$$

$$V_g = V_s$$

$$W_g = W_s \cos \alpha_1 + U_s \sin \alpha_1$$

$$A_{T_{x_g}} = A_{T_{x_s}} \cos \alpha_1 - A_{T_{z_s}} \sin \alpha_1$$

$$A_{T_{y_g}} = A_{T_{y_s}}$$

$$A_{T_{z_g}} = A_{T_{z_s}} \cos \alpha_1 + A_{T_{x_s}} \sin \alpha_1$$

Finally, using the perturbation technique we introduce the relation

$$\lambda_{E1} = L_{E1} + \beta'_{E1}$$

$$\alpha = \alpha_1 + \alpha'$$

$$A_{T_{x_s}} = A_{T_{x_1}} + a_{tx} \text{ etc.}$$

Introducing these perturbations and assuming that V_f , T_f , and J_f are constant for this analysis we have the following steady state relations

$$\begin{aligned}
 V_{fE1}^L = & B_{E1} [(U_{T1} - U_1) \cos \alpha_1 - (W_{T1} - W_1) \sin \alpha_1] \\
 & - [(W_{T1} - W_1) \cos \alpha_1 + (U_{T1} - U_1) \sin \alpha_1] \\
 & - J_V V_{A1} \alpha_1 - \frac{T_f}{2} [(A_{T_{Z1}} \cos \alpha_1 + A_{T_{X1}} \sin \alpha_1] \\
 & - B_{E1} (A_{T_{X1}} \cos \alpha_1 - A_{T_{Z1}} \sin \alpha_1]
 \end{aligned} \tag{c}$$

and

$$\begin{aligned}
 -V_{fAz}^L = & (V_{T1} - V_1) - B_{Az} [(U_{T1} - U_1) \cos \alpha_1 - (W_{T1} - W_1) \sin \alpha_1] \\
 & + J_V V_{A1} \beta_1 + \frac{T_f}{2} [A_{T_{Y1}} - B_{Az} (A_{T_{X1}} \cos \alpha_1 - A_{T_{Z1}} \sin \alpha_1)]
 \end{aligned}$$

Furthermore the perturbation equations are

$$\begin{aligned}
 V_{fE1}^{\lambda} = & B_{E1} [(u_T - u) \cos \alpha_1 - (w_T - w) \sin \alpha_1] \\
 & + \beta_{E1}' [(U_{T1} - U_1) \cos \alpha_1 - (W_{T1} - W_1) \sin \alpha_1] \\
 & - [(w_T - w) \cos \alpha_1 + (u_T - u) \sin \alpha_1] \\
 & - \frac{T_f}{2} (a_{T_z} \cos \alpha_1 + a_{T_x} \sin \alpha_1) \\
 & - B_{E1} (a_{T_x} \cos \alpha_1 - a_{T_z} \sin \alpha_1) \\
 & - \beta_{E1}' (A_{T_{X1}} \cos \alpha_1 - A_{T_{Z1}} \sin \alpha_1) \\
 & - J_V V_{A1} \alpha'
 \end{aligned} \tag{d}$$

and

$$\begin{aligned}
 -V_{fAz}^{\lambda} = & (v_T - v) - B_{Az} [(u_T - u) \cos \alpha_1 - (w_T - w) \sin \alpha_1] \\
 & - \beta_{Az}' [(U_{T1} - U_1) \cos \alpha_1 - (W_{T1} - W_1) \sin \alpha_1] \\
 & + \frac{T_f}{2} [a_{T_y} - B_{Az} (a_{T_x} \cos \alpha_1 - a_{T_z} \sin \alpha_1)] \\
 & - \beta_{Az}' (A_{T_{X1}} \cos \alpha_1 - A_{T_{Z1}} \sin \alpha_1) \\
 & + J_V V_{A1} \beta'
 \end{aligned}$$

B.2 LCOS Equation

Returning to equation A in the previous section, recall that the perfect director assumes perfect knowledge of the line-of-sight rate and target acceleration. The LCOS system, however, uses attacker rotation rates and acceleration to approximate the above parameters.

In terms of the line of sight rate in the attacker's gun coordinate system, $\frac{\dot{\alpha}}{\beta_{A1}}$ we have from Ref. 5, page 6-35

$$\dot{\bar{\beta}} \approx \bar{\omega}_g - \dot{\bar{\lambda}}$$

In this case,

$$\begin{aligned}\dot{\bar{\beta}} - (\dot{\bar{\beta}} \cdot \bar{s})\bar{s} &= (\bar{\omega}_g - \dot{\bar{\lambda}}) - [(\bar{\omega}_g - \dot{\bar{\lambda}}) \cdot \bar{s}]\bar{s} \\ &= (\bar{\omega}_g - \dot{\bar{\lambda}}) - (\bar{\omega}_g \cdot \bar{s})\bar{s} + (\dot{\bar{\lambda}} \cdot \bar{s})\bar{s}\end{aligned}$$

In addition, the attacker acceleration \bar{A}_A is substituted in place of the target's, resulting in the vector equation for the LCOS

$$V_f \bar{\lambda} = D[(\bar{\omega}_g - \dot{\bar{\lambda}}) - (\bar{\omega}_g \cdot \bar{s})\bar{s}] + J_V V_A \bar{\alpha}_{gun} + (\bar{s} \times \bar{A}_A) \frac{T_f}{2}$$

with

$$\bar{s} = i_g + \lambda_{Az} \bar{j}_g + \lambda_{EL} \bar{k}_g$$

$$\bar{\omega}_g = P_g \bar{i}_g + Q_g \bar{j}_g + R_g \bar{k}_g$$

and

$$\bar{A}_A = A_x \bar{i}_g + A_y \bar{j}_g + A_z \bar{k}_g$$

we have the following two scalar differential equations

$$\begin{aligned}D\dot{\lambda}_{EL} &= -V_f \lambda_{EL} + DQ_g - D(P_g + \lambda_{Az} Q_g + \lambda_{EL} R_g) \lambda_3 \\ &\quad - J_V V_A \alpha + \frac{T_f}{2} (\lambda_{EL} A_x - A_z)\end{aligned}\tag{F}$$

and

$$\begin{aligned}-D\dot{\lambda}_{Az} &= V_f \lambda_{Az} + DR_g - D(P_g + \lambda_{Az} Q_g + \lambda_{EL} R_g) \lambda_{EL} \\ &\quad + J_V V_A \beta + \frac{T_f}{2} (A_y - \lambda_{Az} A_x)\end{aligned}$$

The steady state equations are of course

$$\begin{aligned}D_1 \dot{L}_{EL} = 0 &= -V_f L_{Az} + D_1 Q_1 - J_V V_A \alpha_1 \\ &\quad + \frac{T_f}{2} [L_{EL} (A_{x_1} \cos \alpha_1 - A_{z_1} \sin \alpha_1) - (A_{z_1} \cos \alpha_1 + A_{x_1} \sin \alpha_1)] \\ &\quad - D_1 L_{Az} [(P_1 \cos \alpha_1 - R_1 \sin \alpha_1) + L_{Az} Q_1 + L_{EL} (R_1 \cos \alpha_1 + \\ &\quad P_1 \sin \alpha_1)]\end{aligned}$$

$$\begin{aligned}-D_1 \dot{L}_{Az} = 0 &= V_f L_{Az} + D_1 (R_1 \cos \alpha_1 + P_1 \sin \alpha_1) \\ &\quad + J_V V_A \beta_1 + \frac{T_f}{2} [A_{y_1} - L_{Az} (A_{x_1} \cos \alpha_1 - A_{z_1} \sin \alpha_1)] \\ &\quad - D_1 L_{EL} [(P_1 \cos \alpha_1 - R_1 \sin \alpha_1) + L_{Az} Q_1 + L_{EL} (R_1 \cos \alpha_1 + \\ &\quad P_1 \sin \alpha_1)]\end{aligned}\tag{G}$$

Finally, the perturbation equations become

$$\begin{aligned}
 D_1 \lambda'_{\dot{E}L} = & -V_g \lambda'_{\dot{E}1} + D_1 q + Q_1 d - J_V V_A \alpha' \\
 & + \frac{T_f}{2} [L_{EL} (a_x \cos \alpha_1 - a_z \sin \alpha_1) \\
 & + \lambda'_{\dot{E}1} (A_{x1} \cos \alpha_1 - A_{z1} \sin \alpha_1) \\
 & - (a_z \cos \alpha_1 + a_x \sin \alpha_1)] \\
 & - dL_{AZ} M_1 - D_1 \lambda'_{\dot{A}Z} M_1 \\
 & - D_1 L_{AZ} [(p \cos \alpha_1 - r \sin \alpha_1) + \lambda'_{\dot{A}Z} Q_1 + L_{AZ} q \\
 & + \lambda'_{\dot{E}1} (R_1 \cos \alpha_1 + P_1 \sin \alpha_1) \\
 & + L_{EL} (r \cos \alpha_1 + p \sin \alpha_1)] \quad (H)
 \end{aligned}$$

$$\begin{aligned}
 -D_1 \lambda'_{\dot{A}Z} = & V_f \lambda'_{\dot{A}Z} + D_1 (r \cos \alpha_1 + p \sin \alpha_1) \\
 & + d(R_1 \cos \alpha_1 + P_1 \sin \alpha_1) + J_V V_A \beta' \\
 & + \frac{T_f}{2} [a_y - \lambda'_{\dot{A}Z} (A_{x1} \cos \alpha_1 - A_{z1} \sin \alpha_1) \\
 & - L_{AZ} (a_x \cos \alpha_1 - a_z \sin \alpha_1)] \\
 & - dL_{EL} M_1 - D_1 \lambda'_{\dot{E}1} M_1 \\
 & - D_1 L_{EL} [(p \cos \alpha_1 - r \sin \alpha_1) + L_{AZ} q + \lambda'_{\dot{A}Z} Q_1 \\
 & + \lambda'_{\dot{E}1} (R_1 \cos \alpha_1 + P_1 \sin \alpha_1) + L_{EL} (r \cos \alpha_1 + p \sin \alpha_1)]
 \end{aligned}$$

where in the above relations

$$\begin{aligned}
 M_1 = & (P_1 \cos \alpha_1 - R_1 \sin \alpha_1) + L_{AZ} Q_1 \\
 & + L_{EL} (R_1 \cos \alpha_1 + P_1 \sin \alpha_1)
 \end{aligned}$$

C. Target Kinematics

In the foregoing sections, the equations for the line-of-sight rates and lead angles were derived in terms of the targets velocities and accelerations in the attacker's stability coordinate system. In this section the relations governing these velocities and accelerations will be developed.

Now, in a moving attacker coordinate system with

$$\bar{V}_{\text{Target}} = U_T \bar{i}_A + V_T \bar{j}_A + W_T \bar{k}_A$$

we have in the stability axis of the attacker

$$\bar{V}_T = U_{T_s} \bar{i}_s + V_{T_s} \bar{j}_s + W_{T_s} \bar{k}_s + \omega_s \times \bar{V}_T$$

Also, since $\dot{\bar{V}}_T$ is the target's acceleration vector, we can write it in terms of accelerations in the stability axis as

$$\dot{\bar{V}}_T = A_{T_{x_s}} \bar{i}_s + A_{T_{y_s}} \bar{j}_s + A_{T_{z_s}} \bar{k}_s$$

Therefore, we have the scalar equations for the target's velocities in terms of accelerations, or

$$\dot{U}_{T_s} = A_{T_{x_s}} + (R_s V_{T_s} - Q_s W_{T_s})$$

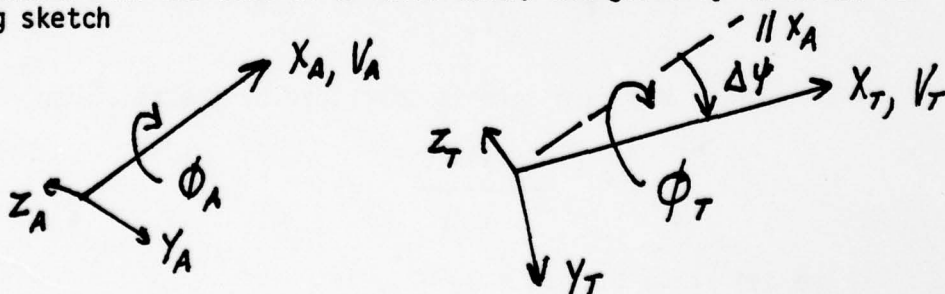
$$\dot{V}_{T_s} = A_{T_{y_s}} + (P_s W_{T_s} - R_s U_{T_s})$$

$$\dot{W}_{T_s} = A_{T_{z_s}} + (Q_s U_{T_s} - P_s V_{T_s})$$

(A)

where P_s , Q_s , R_s are the rotation rates of the attacker's stability axis.

We may now express the target accelerations in terms of the geometry of the engagement. In the case of a level turn, the geometry is shown in the following sketch



X_A and X_T both in level plane

From geometry we can then write the expressions for the target's acceleration in the attacker's coordinates

$$A_{T_{x_A}} = A_{T_{x_T}} \cos \Delta\psi + A_{T_{z_T}} \sin \Delta\psi \sin \phi_T - A_{T_{y_T}} \sin \psi \cos \phi_T$$

$$A_{T_{y_A}} = A_{T_{x_T}} \sin \Delta\psi \cos \phi_A + A_{T_{y_T}} \cos \Delta\psi \cos (\phi_T - \phi_A) - A_{T_{z_T}} \cos \Delta\psi \sin (\phi_T - \phi_A)$$

$$A_{T_{z_A}} = -A_{T_{x_T}} \sin \Delta\psi \sin \phi_A + A_{T_{y_T}} \cos \Delta\psi \sin (\phi_T - \phi_A) + A_{T_{z_T}} \cos \Delta\psi \cos (\phi_T - \phi_A)$$

Now, if it is assumed that in this encounter the target is capable only of significant accelerations in the Z_T direction ($A_{T_{x_T}} = A_{T_{y_T}} = 0$), we have

$$A_{T_{x_A}} = A_{T_{z_T}} \sin \Delta\psi \sin \phi_T$$

$$A_{T_{y_A}} = -A_{T_{z_T}} \cos \Delta\psi \sin (\phi_T - \phi_A)$$

$$A_{T_{z_A}} = A_{T_{z_T}} \cos \Delta\psi \cos (\phi_T - \phi_A)$$

Substituting these into Eqn. A, we have

$$\dot{U}_{T_S} = -A_{T_{z_T}} \sin \Delta\psi \sin \phi_T + (R_S V_{T_S} - Q_S W_{T_S})$$

$$\dot{V}_{T_S} = -A_{T_{z_T}} \cos \Delta\psi \sin (\phi_T - \phi_A) + (P_S W_{T_S} - R_S U_{T_S})$$

$$\dot{W}_{T_S} = A_{T_{z_T}} \cos \Delta\psi \cos (\phi_T - \phi_A) + (Q_S U_{T_S} - P_S V_{T_S})$$

(B)

Since the turn rate is described by the relation

$$\dot{\psi} = \frac{-A_Z \sin \phi}{V}$$

we can state that $\dot{\Delta\psi} = \dot{\psi}_T - \dot{\psi}_A$ is

$$\dot{\Delta\psi} = \frac{-A_{T_{z_T}} \sin \phi_T}{V_T} + \frac{A_{A_z} \sin \phi_A}{V_A}$$

(c)

or, approximating A_{Az} with $Z_\alpha \alpha_A$ we have

$$\dot{\Delta\psi} = \frac{-A_{Tz_T} \sin \phi_T}{V_T} + \frac{Z_\alpha \alpha_A \sin \phi_A}{V_A} \quad (D)$$

The perturbation equation for Eqns B and D may be derived using the relations $a \ll A$

$$\sin(A + a) \approx \sin A + a \cos A$$

$$\cos(A + a) \approx \cos A - a \sin A$$

Introducing the familiar steady state and perturbation variables, we find that the perturbation equations are

$$\begin{aligned} \dot{u}_T &= a_{Tz} \sin \Delta\bar{\psi}_1 \sin \phi_{T1} + \Delta\psi A_{Tz1} \cos \Delta\bar{\psi}_1 \sin \phi_{T1} \\ &\quad + \phi_T A_{Tz1} \sin \Delta\bar{\psi}_1 \cos \phi_{T1} + r V_{Ts1} \\ &\quad + R_1 v_T - q W_{Ts1} - Q_1 \omega_T \\ \dot{v}_T &= -A_{Tz1} \cos \Delta\bar{\psi}_1 (\phi_T - \phi_A) + p W_{Ts1} + P_1 \omega_T \\ &\quad - r U_{Ts1} - R_1 u_T \\ \dot{\omega}_T &= a_{Tz} \cos \Delta\bar{\psi}_1 - A_{Tz1} \sin \Delta\bar{\psi}_1 \Delta\psi + q U_{Ts1} \\ &\quad + Q_1 u_T - p V_{Ts1} - P_1 v_T \end{aligned} \quad (E)$$

And, assuming $A_{Tz1} = A_{Az1}$, $\phi_{T1} = \phi_{A1}$, $V_T = V_A$ we have

$$\dot{\Delta\psi}_{\text{pert}} = -\frac{1}{V_A} [A_{z1} \cos \phi_1 (\phi_T - \phi_A) + \sin \phi_1 (a_{Tz} - Z_\alpha \alpha)] \quad (F)$$

Finally, to model the target's acceleration, a_{Tz} , and bank angle, ϕ_T , one approach is to assume a Markov process in the form

$$\dot{n}_1 = -\frac{1}{\tau_a} n_1 + \xi_1$$

$$\dot{n}_2 = -\frac{1}{\tau_\phi} n_2 + \xi_2$$

$$\dot{a}_{T_z} = -\frac{1}{\tau_a} a_{T_z} + n_1$$

$$\dot{\phi}_T = -\frac{1}{\tau_\phi} \phi_T + n_2$$

(G)

where ξ_1 and ξ_2 are white noise processes with variances selected from the relations

$$\sigma_{\xi_1}^2 = \frac{4 \sigma_{a_t}^2}{\tau_a^3}$$

$$\sigma_{\xi_2}^2 = \frac{4 \sigma_{\phi_t}^2}{\tau_\phi^3}$$

Consequently, selecting the target time constants, τ_a and τ_ϕ , and the desired variances on acceleration and bank angle, $\sigma_{a_t}^2$ and $\sigma_{\phi_t}^2$, completely describes the target motions with Eqns E, F, and G.

D. F106 Perturbation Equations

In this section, the linearized equations for the vehicle dynamics are summarized. From Reference 8, the perturbation equations for a general steady-state flight condition are as follows:

$$\begin{aligned}
 \dot{u} - V_1 r - R_1 v + W_1 q + Q_1 w &= -g\theta \cos\theta_1 + X_u u + X_w w + X_{\delta_E} \delta_E \\
 \dot{v} + U_1 r + R_1 u - W_1 p - P_1 w &= -g\theta \sin\theta_1 \sin\theta_1 \\
 &\quad + g\phi \cos\phi_1 \cos\theta_1 + Y_v v + Y_p p \\
 &\quad + Y_r r + Y_{\delta_A} \delta_A + Y_{\delta_R} \delta_R \\
 \dot{w} - U_1 q - Q_1 u + V_1 p + P_1 v &= -g\theta \cos\phi_1 \sin\theta_1 \\
 &\quad - g\phi \sin\phi_1 \cos\theta_1 + Z_u u \\
 &\quad + Z_w w + Z_{\delta_E} \delta_E \\
 \dot{q} + \frac{1}{I_{yy}} (I_{xx} - I_{zz}) (P_1 r + R_1 p) + I_{xz}/I_{yy} (2P_1 p - 2R_1 r) \\
 &= M_u u + M_w w + M_{\dot{w}} \dot{w} + M_q q + M_{\delta_E} \delta_E \\
 \dot{p} - I_{xz}/I_{xx} \dot{r} - I_{xz}/I_{xx} (P_1 q + Q_1 p) \\
 &\quad + \frac{1}{I_{xx}} (I_{zz} - I_{yy}) (R_1 q + Q_1 r) = L_{\beta} \beta + L_p p \\
 &\quad + L_r r + L_{\delta_A} \delta_A + L_{\delta_R} \delta_R \\
 \dot{r} - I_{xz}/I_{zz} \dot{p} + \frac{I_{xz}}{I_{zz}} (Q_1 r + R_1 q) + \frac{1}{I_{zz}} (I_{yy} - I_{xx}) (P_1 q + Q_1 p) \\
 &= N_{\beta} \beta + N_p p + N_r r + N_{\delta_A} \delta_A + N_{\delta_R} \delta_R
 \end{aligned} \tag{A}$$

where in these relations the $(\cdot)_1$ quantities (e.g., U_1, V_1, W_1) are constant, steady-state values, and X_w, Z_w , etc are the dimensional stability derivatives (see the above reference).

The case to be evaluated consists of a level, 4-g turning condition at 10,000 ft altitude, Mach = 0.72. Therefore, for this case, the steady-state parameters are given as follows (equations developed in the aircraft stability axis)

$$\theta_1 = 0.$$

$$U_1 = V = 775 \text{ ft/sec}$$

$$\cos \phi_1 = \frac{1}{n} = \frac{1}{4}$$

$$V_1 = W_1 = 0.$$

$$\phi_1 = 75.52 \text{ deg.}$$

$$\text{Turn rate, } \dot{\psi} = g (\tan \phi_1) / U_1 = .161 \text{ rad/sec.}$$

$$\text{Roll rate, } P_1 = 0$$

$$\text{Pitch rate, } Q_1 = \dot{\psi} \sin \phi_1 = .156 \text{ rad/sec}$$

$$\text{Yaw rate, } R_1 = \dot{\psi} \cos \phi_1 = .040 \text{ rad/sec.}$$

(B)

The angle of attack required (Ref. 9) is $\approx 14.5^\circ$ and the elevator deflection for trim is -20° at this altitude and velocity. The inertias (in the body axis) are

$$I_{xx} = 2.126 \times 10^4 \text{ slug-ft}^2$$

$$I_{yy} = 2.035 \times 10^5 \text{ slug-ft}^2$$

$$I_{zz} = 2.175 \times 10^5 \text{ slug-ft}^2$$

$$I_{xz} = 7.316 \times 10^3 \text{ slug-ft}^2$$

The reference wing area is 695 ft^2 , and the weight is 33,000 pounds.

In the stability axis, the inertias are

$$I_{xx_s} = 2.999 \times 10^4 \text{ slug-ft}^2$$

$$I_{zz_s} = 2.088 \times 10^5 \text{ slug-ft}^2$$

$$I_{xz_s} = -4.111 \times 10^4 \text{ slug-ft}^2$$

Finally, the dimensional derivatives are estimated to be (for $c = \text{m.a.c.} = 23.8 \text{ ft}$)

$$X_u = -.0924$$

$$X_\alpha = -176.0$$

$$X_\delta = -24.4$$

$$Z_u^{\delta E} = -.437$$

$$Z_\alpha = -987.$$

$$Z_\delta = -288.$$

$$M_u^{\delta E} = -.00223$$

$$M_\alpha = -11.5$$

$$M_\alpha = -.785(.3485)$$

$$M_q = -.785(.3485)$$

$$M_{\delta E} = -17.36$$

$$Y_\beta = -177.$$

$$Y_p = 1.76$$

$$Y_r = 8.13$$

$$Y_\delta = 84.9$$

$$Y_\delta^A = 36.5$$

$$L_\beta^r = -47.6$$

$$L_p = -.0613$$

$$L_r = 2.77$$

$$L_{\delta A} = -57.4$$

$$L_{\delta R} = .567$$

$$\begin{aligned} N_{\beta} &= 6.77 \\ N_p &= -.176 \\ N_r &= -.842 \end{aligned}$$

$$\begin{aligned} N_{\delta_A} &= -5.93 \\ N_{\delta_R} &= -3.98 \end{aligned}$$

Finally, in the steady-state condition being considered, the aileron/rudder interconnect results in the following relation

$$\delta_R = \delta_R \text{ commanded} - 1.1\delta_A$$

And, in addition, the feel system for this vehicle is estimated to provide the following stick gains

$$\begin{aligned} \delta_E &= .0467 \delta_{E_{stick}} \text{ (in.)} \\ \delta_A &= .0273 \delta_{A_{stick}} \text{ (in.)} \\ \delta_R &= .0738 \delta_{R_{ped.}} \text{ (in.)} \end{aligned}$$

Combining all above parameters, the resulting vehicle dynamics are

$$\begin{aligned} \dot{u} &= -.0924 u - 297\alpha - 32.2\theta + 31.2\beta - 1.14\delta_{E_{st}} \\ \dot{\alpha} &= -.00036 u - 1.272\alpha + q - .040\phi - .0173\delta_{E_{st}} \\ \dot{q} &= -.00013 u - 11.21\alpha - .547q + .0109\phi \\ &\quad + .0353 p - .0162r - .806\delta_{E_{stick}} \\ \dot{\beta} &= -.228\beta + .0104\phi + .00227p - .99r \\ &\quad - .00005u + .00298\delta_{A_{st}} + .00347\delta_{R_{ped}} \\ \dot{p} &= -.0245q - 77.9\beta + .198p + 5.27r \\ &\quad - 1.843\delta_{A_{st}} + .608\delta_{R_{st}} \\ \dot{r} &= .0127q + 22.1\beta - .344p - 1.852r \\ &\quad + .201\delta_{A_{st}} - .413\delta_{R_{st}} \end{aligned} \tag{c}$$

In addition to the vehicle dynamic model, the vehicle kinematics are described by the perturbation relations

$$\begin{aligned} p &= \dot{\phi} - \dot{\psi}_1 \theta \\ q &= \dot{\theta} \cos\phi_1 + \dot{\psi}_1 \phi \cos\phi_1 + \dot{\psi} \sin\phi_1 \\ r &= -\dot{\psi}_1 \phi \sin\phi_1 + \dot{\psi} \cos\phi_1 - \dot{\theta} \sin\phi_1 \end{aligned}$$

In our case, solving for $\dot{\theta}$, $\dot{\phi}$, and $\dot{\psi}$ yields

$$\dot{\phi} = p + .161\dot{\theta}$$

$$\dot{\theta} = -.161\dot{\phi} + .25q - .968r$$

$$\dot{\psi} = .968q + .25r$$

(D)

Equations C and D now constitute the final vehicle model.

PROGRAM PIREP: LOSS CASE D=3000FT F-4E AT=3.5 GS +++++AUGMNTD

NO. OF STATES 9
 NOISE SHAPING STATES 3
 NO. OF CONTROLS 1
 NO. OF NOISE SOURCES 1
 NO. OF OUTPUTS 4
 KTRG 2

E. Sample Output

SYSTEM DYNAMICS ARE: $\dot{X} = AX + BU + EW$, $Y = CX + DU$

A MATRIX:

-0.3333	0.	0.	0.	0.	0.
1.000	-0.3333	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.
0.	-1.0500E-03	-1.0000E-02	0.	0.	0.
0.	0.	0.	0.	0.	0.
0.	0.	.3172	0.	0.	.3172
0.	-0.3172	0.	0.	0.	0.
0.	0.	0.	0.	-0.7660	.3500
0.	0.	1.000	0.	0.	0.
0.	0.	0.	0.	0.	-1.033
0.	0.	1.000	-9.5120E-02	0.	0.
0.	0.	0.	0.	0.	0.
0.	0.	1.000	0.	0.	-10.09
0.	0.	0.	0.	0.	0.
0.	0.	-1.082	-37.05	0.	0.
4.7680E-04	1.2560E-03	-1.840	-3.936	-1.520	-3.104
	5.792	1.984	-23.01		

OPEN-LOOP EIGENVALUES:

-19.18	0.	J
-2.725	2.942	J
-2.725	-2.942	J
-0.1000E-01	0.	J
-0.3333	0.	J
-0.3333	0.	J
-0.3723	0.	J
-0.4450	.3356	J
-0.4450	-0.3356	J

B MATRIX:

0.
 0.
 0.
 0.
 0.
 0.
 0.
 0.
 12.61

C MATRIX:

0.	0.	0.	1.000	1.000	0.
0.	-1.000	0.	0.	0.	0.
0.	0.	.3172	0.	-0.7660	.6600
0.	-0.3172	0.	0.	0.	0.
0.	0.	0.	0.	1.000	0.
0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	-0.7660	.3500
0.	0.	1.000	0.	0.	0.

THIS PAGE IS BEST QUALITY PRACTICABLE
 FROM COPY FURNISHED TO DDO

D MATRIX:

PROGRAM PIREP: LOSS CASE D=3000FT F-4E AT=3.5 GS +++++AUGMENTED

52

0.
0.
0.
0.

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO

F MATRIX:
.3849

0.
0.
0.
0.
0.
0.
0.
0.

COST FUNCTIONAL WEIGHTINGS

STATE

0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.

OUTPUT

16.00	1.000	0.	4.000
-------	-------	----	-------

CONTROL

GT.RATE

.2467

GRAMMIAN IS 7X 7 OF RANK 6

RICCATI SOLN IN 11 ITERATIONS

FEEDBACK CONTROL IS $TN \cdot UDOT + U = -L OPT \cdot X$, WHERE OPTIMAL GAINS (LOPT):
-7.3115E-05 -1.7494E-04 .2443 .4947 .1786 .5575
-.7407 -.2244 .3961

TN MATRIX:
.1001

EIGENVALUES:
.1001

0. J

FEEDBACK CONTROL IS ALSO $UDOT = -LX(X) - LU(U)$ WHERE OPTIMAL GAINS, LX, LU:
-7.3035E-04 -1.7475E-03 2.440 4.941 1.784 5.569
-7.399 -2.242 3.957 9.989

CLOSED-LOOP EIGENVALUES:

-19.77	0.	J
-6.447	7.419	J
-6.447	-7.419	J
-.7439	.8723	J
-.7439	-.8723	J
-.1000E-01	0.	J
-.2476	0.	J
-.3333	0.	J
-.3333	0.	J
-1.482	0.	J

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO

CONTROLLER TIME DELAY: .200

VARIANCE OF RANDOM TURBULENCE:
1.2701E+04MOTOR NOISE: (RATIOS IN DB)
-30.00OBSERVATIONAL THRESHOLDS:
1.1340E-02 2.3100E-02 1.1340E-02 2.3000E-02SENSOR NOISE: (RATIOS IN DB)
-20.00 -20.00 -20.00 -20.00ATTENTIONAL ALLOCATION:
1.000 1.000 1.000 1.000

LIN EQN ALGORITHM NON-CONVERGENT 12 ITERATIONS

RICCATI BLOW UP AT ITERATION 2 INITIAL T= 1.38939

RESET WITH T= 2.77878

RICCATI SOLN IN 17 ITERATIONS

RICCATI SOLN IS PSD--RANK 6

LIN EQN ALGORITHM NON-CONVERGENT 7 ITERATIONS

RICCATI BLOW UP AT ITERATION 2 INITIAL T= 1.38939

PROGRAM PIREP: LOSS PAGE 0=3000FT F-4E AT=3.5 GS +++++AUGMNTD

RESET WITH T= 2.77878

RICCATI SOLN IN 12 ITERATIONS

RICCATI SOLN IN 4 ITERATIONS

RICCATI SOLN IN 3 ITERATIONS

RICCATI SOLN IN 2 ITERATIONS

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDG

URMS AND MOTOR NOISE VARIANCE

3.5869E-02

4.3420E-06

4.3420E-06

VRMS AND NOISE VARIANCE AT ITERATION 51

1.3011E-02 9.1070E-03 .1660 3.7987E-02

3.6157E-05 1.9393E-02 9.6848E-04 1.5267E-04

3.6159E-05 1.9480E-02 9.6847E-04 1.5267E-04

COST GRADIENT WRTO F:

-3.7842E-04 -2.3246E-06 -4.1054E-04 -2.1698E-03

TOTAL COST, J(U)= .9259E-02

SAMPLING COST= .4399E-02

OPTIMAL ESTIMATOR GAINS:

-121.6 -3.400 -197.7 -2214.

-1784. -7.996 -404.0 -2497.

6.452 1.2209E-02 .6648 2.179

2.422 2.2432E-03 .1435 .2149

1.637 3.0272E-03 .2018 .6749

1.091 2.1962E-03 .1432 .5534

3.130 4.3109E-03 .2843 .7425

1.494 3.9469E-03 .2109 1.781

-.3595 -8.7159E-04 -4.6318E-02 -.2521

9.7466E-03 6.9894E-05 -6.0995E-04 -.1096

*** RMS MODEL PREDICTIONS ***

INDEX	X	Y	VY	VYEFF	PY(DB)	FC(%)
1	53.13	.1301E-01	.2306E-02	.6013E-02	-20.0	25.0
2	112.7	.9107E-02	.1610E-02	.1393	-20.0	25.0
3	2.836	.1660	.2943E-01	.3112E-01	-20.0	25.0
4	2.812	.3799E-01	.6733E-02	.1236E-01	-20.0	25.0
5	.1660	0.	0.	0.	0.0	0.0
6	.9599E-01	0.	0.	0.	0.0	0.0
7	2.815	0.	0.	0.	0.0	0.0
8	.1000	0.	0.	0.	0.0	0.0
9	.2888E-01	0.	0.	0.	0.0	0.0

	U	VU	PU(DB)	UDOT
1	.3587E-01	.2010E-02	-30.00	.5314E-01

PROGRAM PREP: LOSS CASE D=3000FT F-4E AT=3.5 GS *****AUGMNTD

COV. MATRIX INCLUDING STATES AND CONTR.

2823.	4235.	-12.95	-5.328	-3.482	-2.271
	-9.199	-3.033	.7343	.7995	
4235.	1.2705E+04	-76.58	-43.61	-15.70	-9.515
	-58.65	-10.45	2.908	3.458	
-12.95	-76.58	8.041	7.909	5.6975E-02	2.7614E-02
	7.964	1.8062E-02	-7.7554E-03	-1.0920E-02	
-5.328	-43.61	7.909	7.906	-2.3225E-03	-6.1849E-03
	7.903	-1.4937E-02	2.3543E-03	1.4722E-03	
-3.482	-15.70	5.6975E-02	-2.3225E-03	2.7562E-02	1.5843E-02
	2.4510E-02	1.5567E-02	-4.7328E-03	-5.8256E-03	
-2.271	-9.515	2.7614E-02	-6.1849E-03	1.5843E-02	9.2140E-03
	9.1648E-03	9.2519E-03	-2.7596E-03	-3.3876E-03	
-9.199	-58.65	7.964	7.903	2.4510E-02	9.1648E-03
	7.923	-8.5033E-14	-2.2254E-03	-4.1852E-03	
-3.033	-10.45	1.8062E-02	-1.4937E-02	1.5567E-02	9.2519E-03
	-8.5033E-14	1.0010E-02	-2.9116E-03	-3.4211E-03	
.7343	2.908	-7.7554E-03	2.3543E-03	-4.7328E-03	-2.7596E-03
	-2.2254E-03	-2.9116E-03	8.3426E-04	1.0229E-03	
.7995	3.458	-1.0920E-02	1.4722E-03	-5.8256E-03	-3.3876E-03
	-4.1852E-03	-3.4211E-03	1.0229E-03	1.2866E-03	

TOTAL ATT'N= 4.00 TOTAL COST= .9259E-02 PERF. COST= .8563E-02
 TIME 14.44.1 DATE 06/30/78

ALL CASES PROCESSED

THIS PAGE IS BEST QUALITY PRACTICABLE
 FROM COPY FURNISHED TO DDO

AIR FORCE AERO-PROPULSION LABORATORY

Research Associates:

Donald R. Jenkins, Lafayette College

Yuen-Koh Kao, University of Cincinnati

Chris C. Lu, University of Dayton

Pau-Chang Lu, University of Nebraska

CATALYTIC FLAME STABILIZATION FOR AIRCRAFT AFTERBURNERS

DONALD R. JENKINS

AUGUST 1978

CATALYTIC FLAME STABILIZATION FOR AIRCRAFT AFTERBURNERS

Donald R. Jenkins

Lafayette College, Easton, Pennsylvania

Abstract

Catalytic flame stabilization utilizes a solid catalytic surface to stabilize and provide a continuous pilot for the flame propagation wave. A test program is being set up to obtain and compare the performance of a J-85-5 turbojet afterburner using a conventional bluff body flame-holder and several different catalyst coated and uncoated ceramic substrate screens for flame-holders. Substrate materials to be used are Cordierite and Silicon Carbide. The catalyst coating's will be platinum and palladium. Least squares mathematical treatment and integration of the data will be used to obtain the overall emission concentrations and the overall combustion efficiency. It is hoped that test results will bear out modeling predictions for an increase in combustion efficiency and decreased pressure drop through the flame-holder. Preliminary work and testing thus far has been in preparation for establishing base line performance. Final test results should establish criteria for selection of both substrate and catalytic materials to be used for flame-holders in afterburners.

August 1978

FOREWARD

This report is to present the progress to date for an on going research project comparing the performance of the J-85-5 afterburner using a conventional bluff body flameholder and several different catalyst coated and uncoated substrate screens for flameholders. This work is being performed as part of the USAF-ASEE summer facility research program in the Fuels Branch, Fuels and Lubrication Division, of the Wright-Patterson Air Force Aero Propulsion Laboratory.

The author would like to give special recognition to Leonard C. Angello for his help, guidance, and supervision in the research work and in the preparation of this report. Recognition and thanks is also extended to Richards Lane and Blaine Heitkamp for their continuous effort as members of the group working on this project.

A special debt of gratitude is owed to: Raymond Allen, Thomas Campbell, Jerrold Carnes, Ralph Malone, Forest Roberts, Joseph Simmons, Wendell Suffron and Robert Whitlock for their invaluable knowledge of the facility and help in preparing for this test.

Appreciation and praise is also extended to Sharon Foley and David Richardson at the Computer Center for their effort and patience in preparing the program to convert the data tape for input to the calculation program.

The dedication and attention to detail on the part of Charlyn Wehner and Elaine Baldwin in preparing this report is gratefully acknowledged.

TABLE OF CONTENTS

	Page
Abstract	
Foreward	
Introduction	1
Apparatus	2
C-Stand Facility	6
Flameholder Design	12
Combustion of a Hydrocarbon Fuel with Air	15
Test Program	20
Data Processing	23
Appendix	
A. Program FAR	25
B. Program DELTA	26
C. ρ Vel Calculations	27
D. Exhaust Nozzle Calibration	30
E. Radial Probe Calibration	32
F. Data for Exhaust Stream Limits	34
G. Hydrogen Check and Velocity Profile	37
H. Input to Data Tape	39
I. Input and Output Format	40
J. Data Summary and Program Output Format	49

LIST OF FIGURES

	Page
1. Major Engine Sections	3
2. Afterburner Assembly	4
3. Diffuser Components	5
4. Engine Control Panel	7
5. Emission Instruments	8
6. Gas Analysis Data Handling	9
7. Exhaust Analysis System for THC and H ₂	10
8. J-85-5 Engine Parameters for Catalytic Flame Stabilization . . .	11
9. J-85-5 Experimental Catalytic Flameholder	13

INTRODUCTION

Various studies have indicated that improvement in the performance of aircraft afterburning engines can be expected through the use of a catalytic flame stabilization device in the afterburner. Flameholder modeling studies have indicated that an increase in efficiency and an improvement in the percent pressure loss can be expected through the use of a catalyst coated substate screen flameholder.

The research test program was to compare results using a conventional V-shaped bluff body flameholder and a honeycomb structure, catalytic flameholding device. The designs to be used were obtained by mathematically modeling a flameholder for a J-85-5 afterburning turbojet engine. The overall purpose of this experimental program is to determine the feasibility of catalytic flameholding devices to stabilize the flame in aircraft afterburners.

To accomplish this purpose and to continue effort on work already done, the following specific objectives were set up:

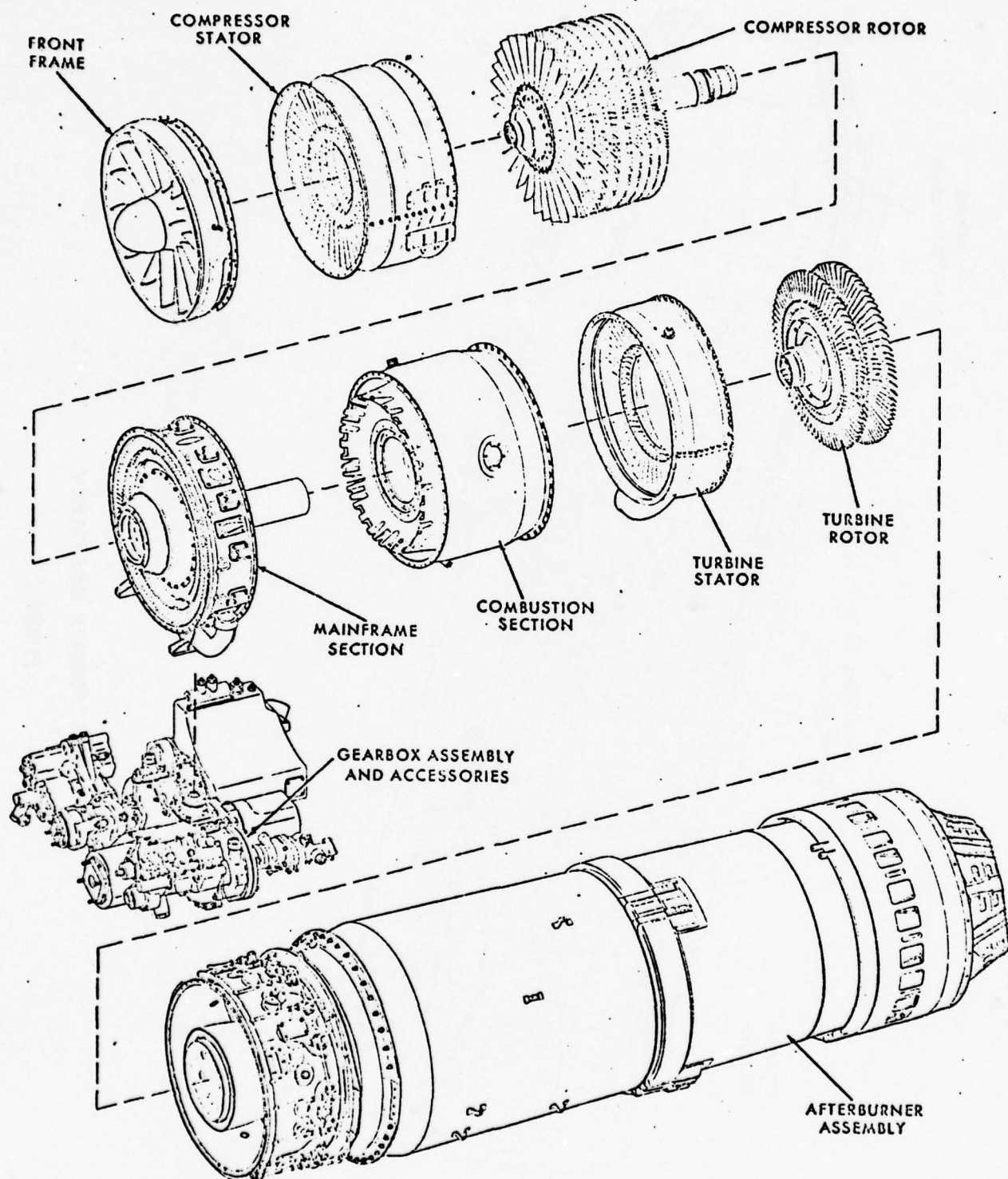
1. To prepare the engine and data acquisition system for actual testing: This has involved instrument trouble-shooting, calibration, and measurement and calibration of the engine nozzle and the radial probe position readout.
2. To set up a test program and procedure. This will be discussed later in this report.
3. To debug, revise, and update the data reduction computer program which will calculate and integrate the data to obtain the overall combustion efficiency and the overall emission concentrations. The program for calculations will be discussed later in this report.

The test program is being performed using a J-85-5 engine at the Air Force Aero Propulsion Laboratory. The work thus far has been mainly preparation for base line and actual testing. The engine has been run three times to establish jet stream diameter limits, to make a hydrogen check and velocity profile, and to check the tape data handling procedure. The next step is to establish base line performance characteristics using the conventional bluff body flameholders. The major performance advantages expected are increased afterburner combustion efficiency and reduced flameholder pressure drop.

APPARATUS

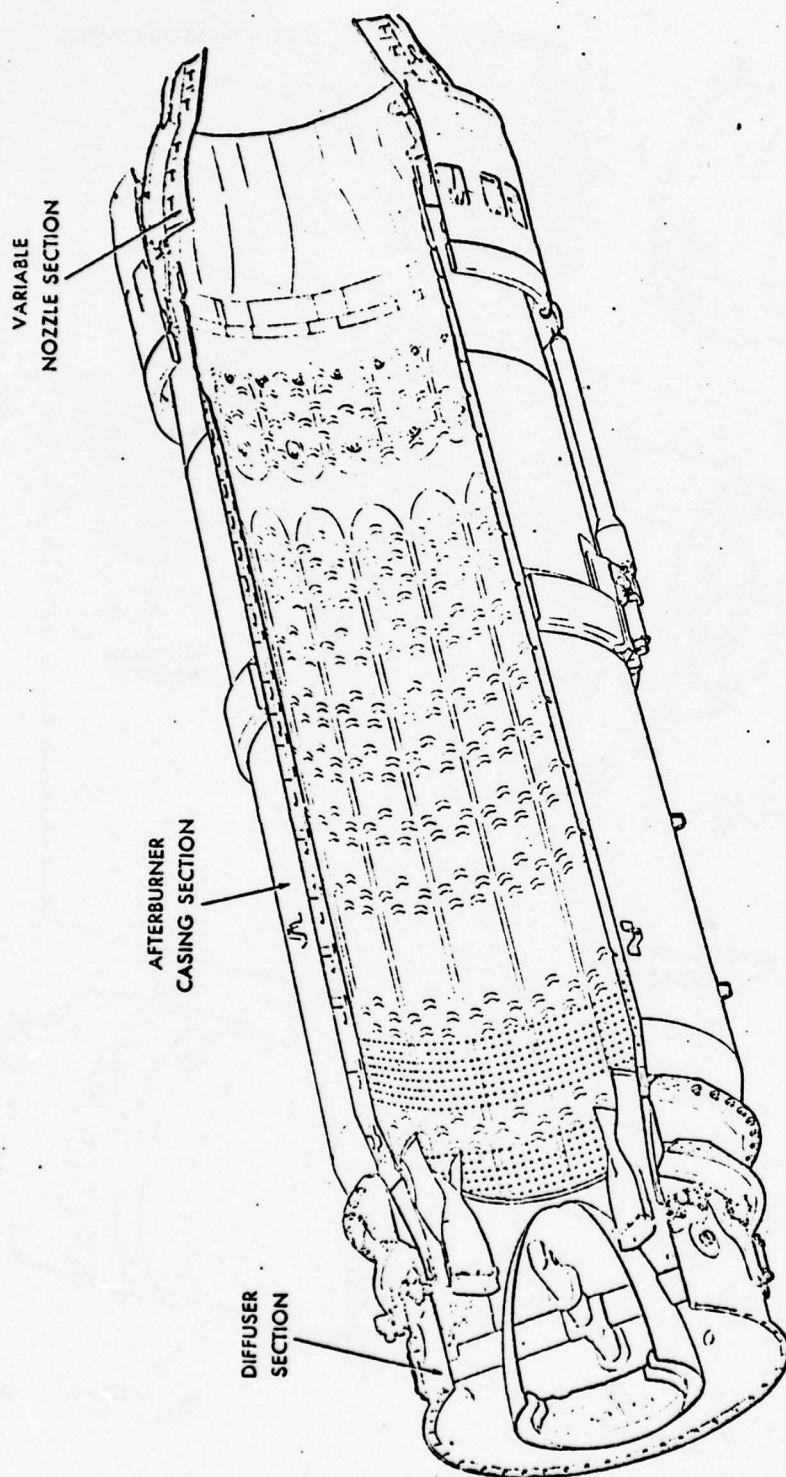
The J-85-5 jet engine has an eight stage compressor, and annular combustion chamber, a two stage turbine and an afterburner. The afterburner consists of a diffuser section housing the flameholders, a cylindrical double walled annular burning section; and finally a variable exhaust nozzle. The major engine sections are shown in Figure 1.

The afterburner consists of a straight through duct equipped with a flameholder needed to stabilize the reacting fuel/air mixture. The afterburner's function is to augment the maximum thrust that can be produced by the main engine. Due to high gas temperatures causing dissociation of H_2O and CO_2 and short residence time, afterburners do not operate at combustion efficiencies as high as in the main burner. It is hoped that through the use of a catalytic flameholding device, faster reaction rates and thus higher efficiencies can be achieved. This may also result in being able to use a shorter length of afterburner. The afterburner assembly is shown in Figure 2. The diffuser components showing the fuel system and the existing bluff body flameholder is shown in Figure 3.



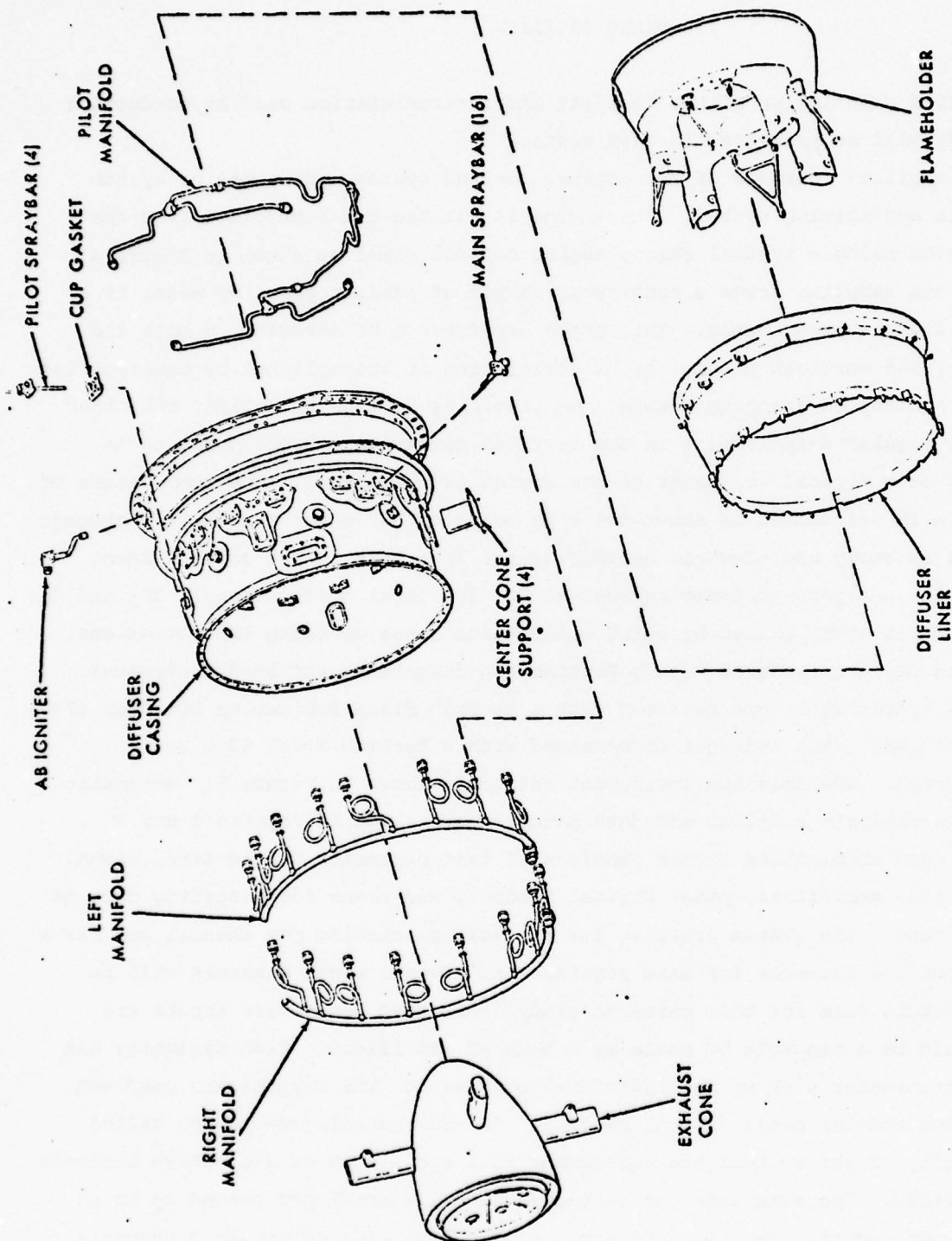
MAJOR ENGINE SECTIONS

FIGURE 1



AFTERBURNER ASSEMBLY

FIGURE 2



DIFFUSER COMPONENTS

FIGURE 3

C-STAND FACILITY

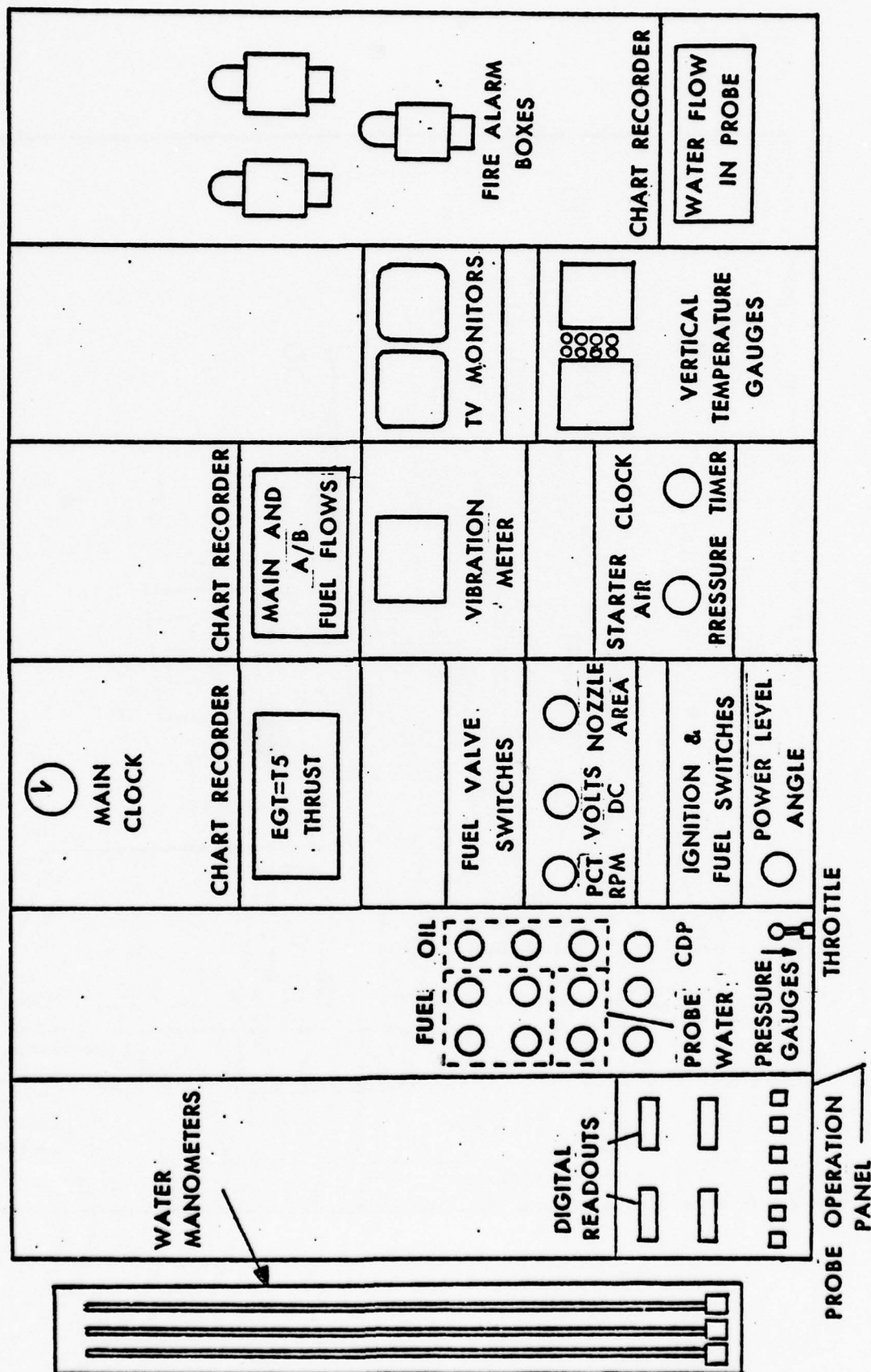
A brief description of the facility and instrumentation used in conducting this study will be presented in this section.

The facility consists of the engine, control system, gas sampling system and a data acquisition system. The engine is started and controlled from the control room using a typical remote engine control panel as shown in Figure 4.

The gas sampling draws a continuous sample of exhaust gases by means of a movable single element probe. This probe has freedom of movement in both the horizontal and vertical plane. Probe positioning is accomplished by means of two remotely controlled stepping motors, one providing horizontal motion, the other providing angular displacement in the vertical direction. Probe position is displayed on a digital voltmeter on the engine control panel. The temperature of the sample is maintained at about 300°F by means of hot water circulating through the probe assembly and electric heating of the line back to the control room.

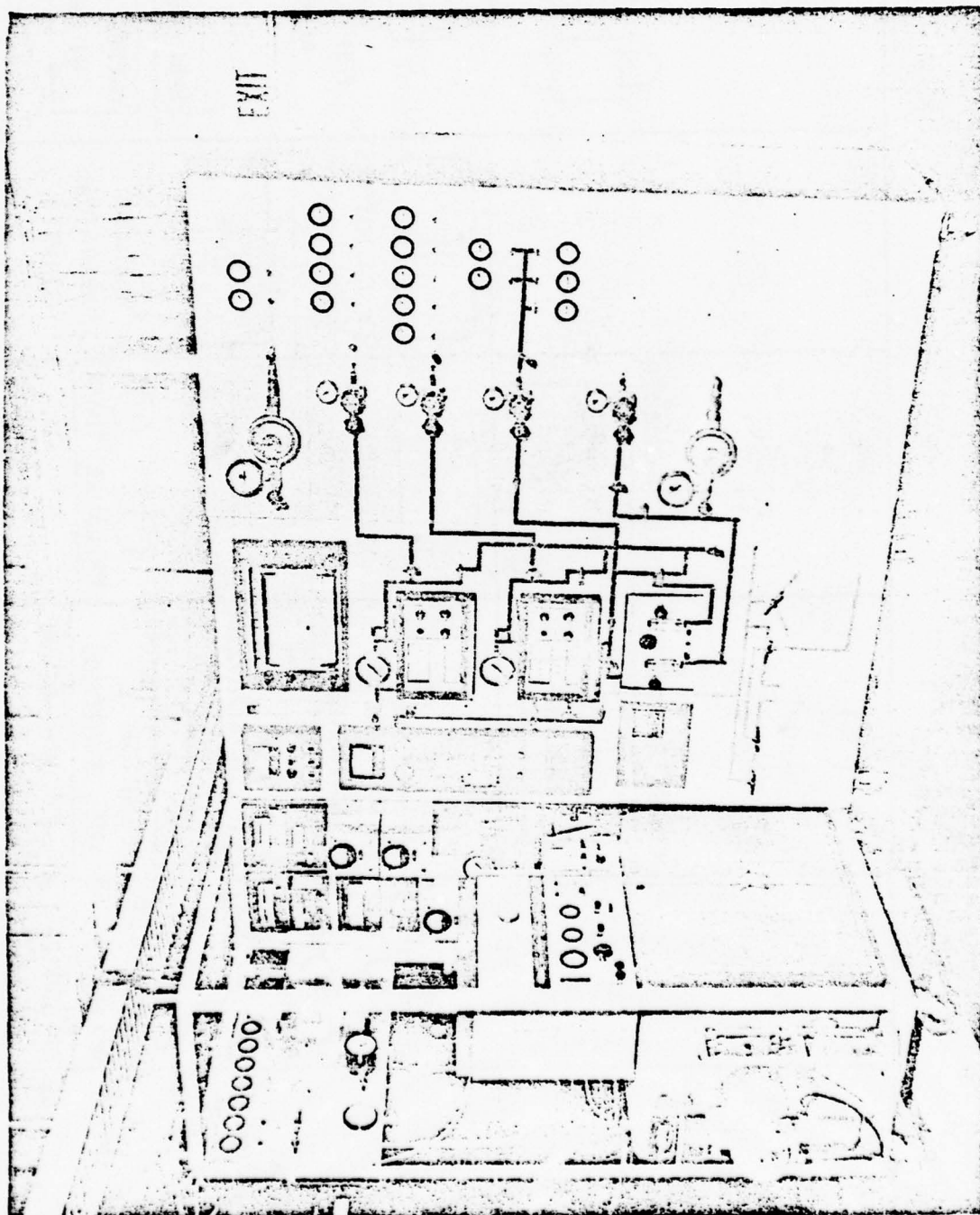
The gas analysis instruments measure CO, THC (total hydrocarbons), CO₂ and H₂. Calibration is accomplished by using calibration gases of known concentrations. The CO and CO₂ are measured with a Beckman non-dispersive infrared instrument. The total hydrocarbons are measured with a Beckman Flame Ionization Detector (FID) type instrument. The hydrogen is measured with a Beckman Model 6700 Gas Chromatograph. The Emission Instrument set up is shown in Figure 5. Schematics of the gas analysis sampling and data handling are shown in Figures 6 and 7.

The data acquisition system consists of test parameter transducers, signal conditioners, amplifiers, panel digital readout, and means for recording data on magnetic tape. The system provides for one voltage reading per channel and has a capacity of 100 channels for data acquisition. Twenty eight channels will be used to obtain data for this research study. All test parameters inputs are conditioned to a ten volt DC scale by a bank of amplifiers. Each parameter has its own transducer pick up and individual amplifier. The signals are read out on a common modular panel digital readout. The 28 channels may be called individually or set so that the instrument will run a scan on all active channels consecutively. The scan rate can be varied from 1 channel per second up to 15 channels per second. For this project the scan rate will be set at 2 channels per second. A schematic of the data acquisition system is shown in Figure 8.



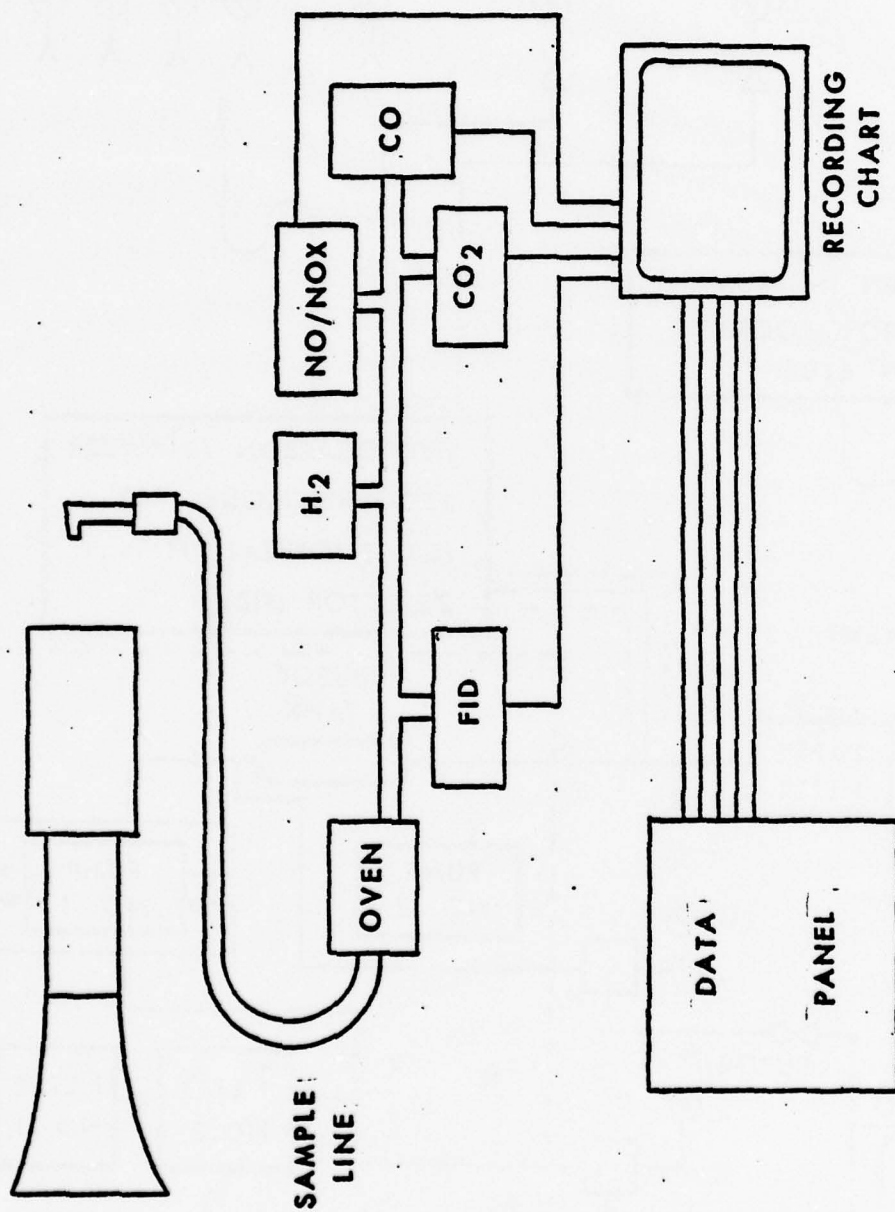
ENGINE CONTROL PANEL

FIGURE 4



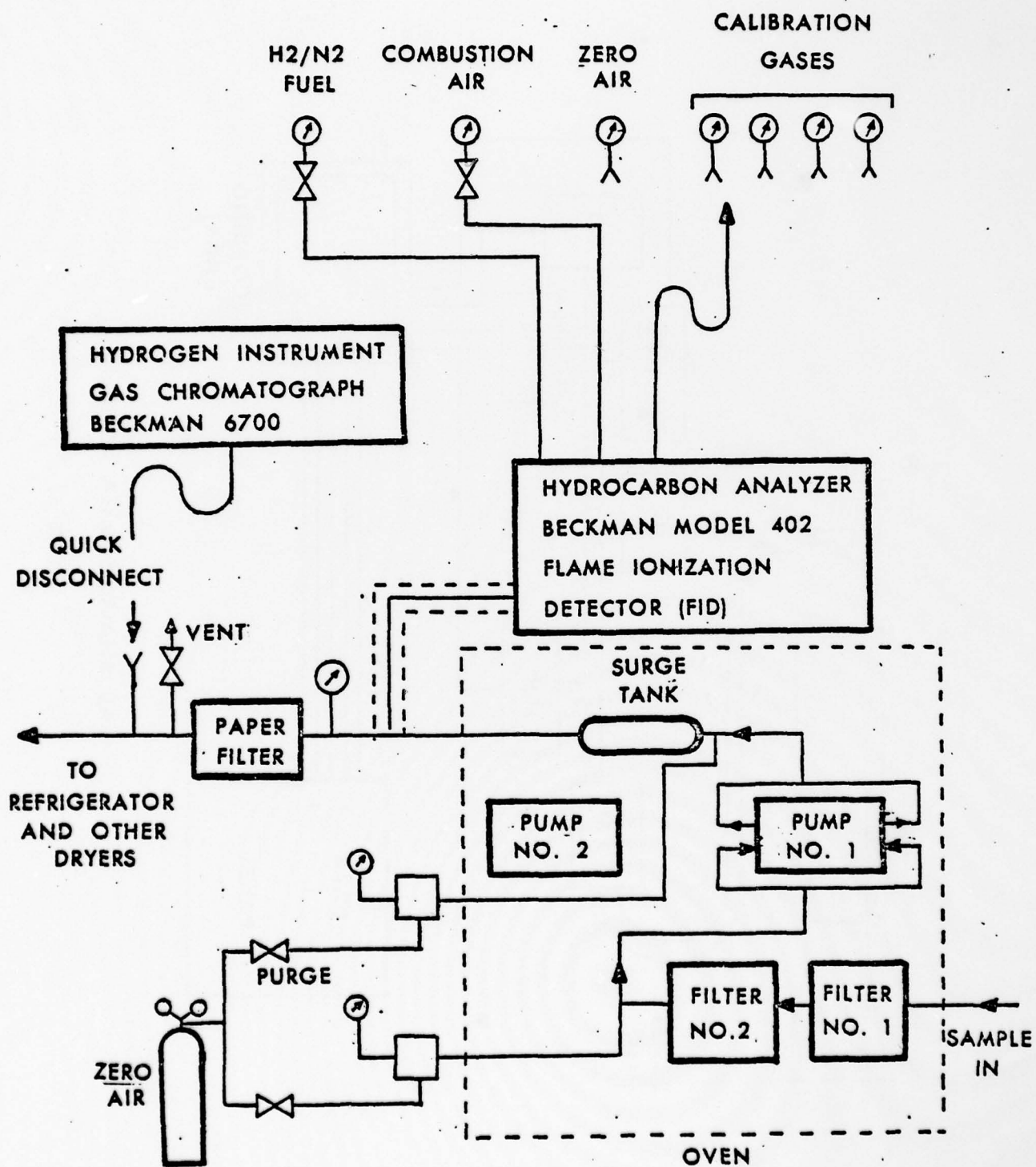
EMISSION INSTRUMENTS

FIGURE 5



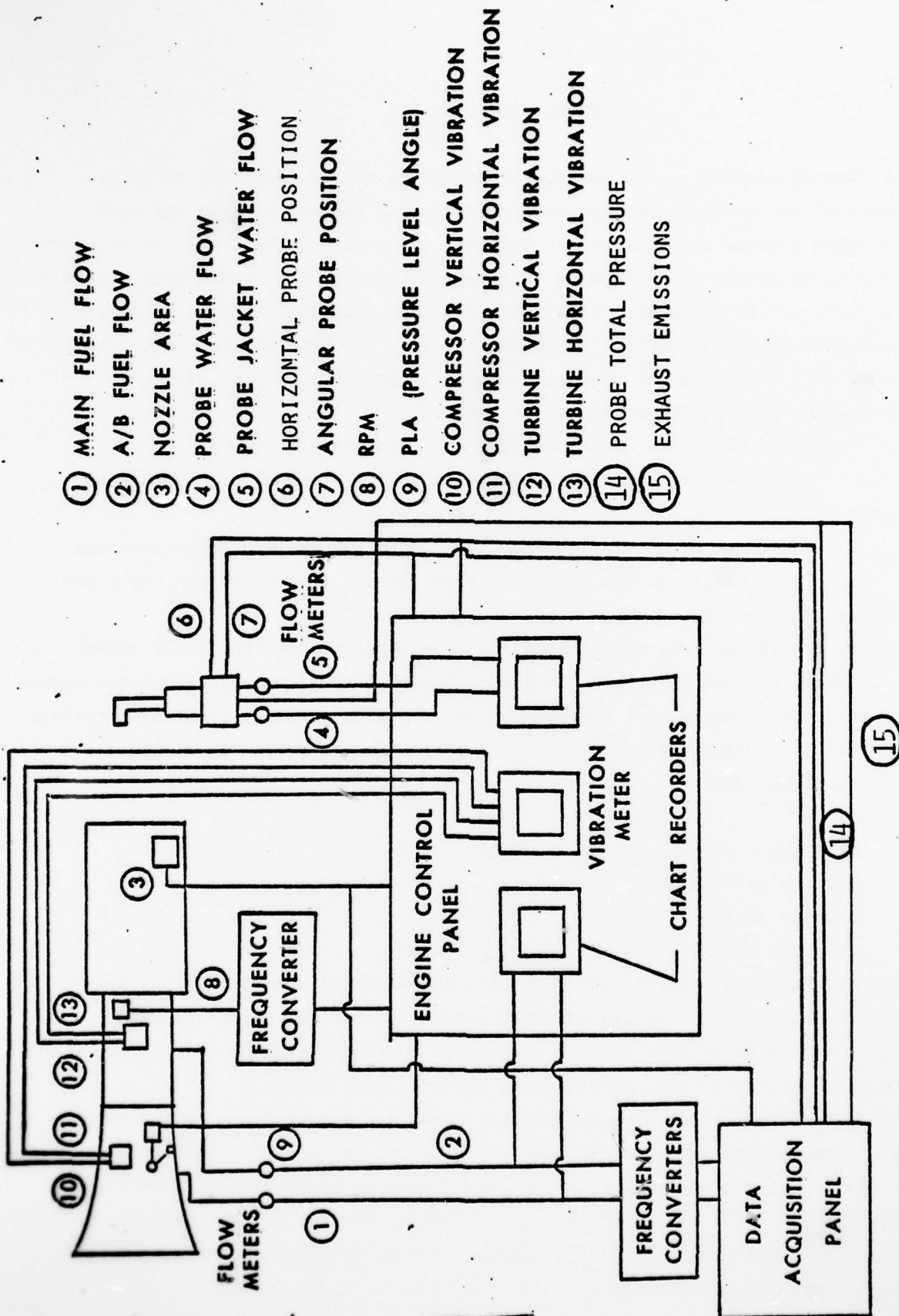
GAS ANALYSIS DATA HANDLING

FIGURE 6



EXHAUST ANALYSIS SYSTEM FOR THC AND H₂

FIGURE 7



J-85-5 ENGINE PARAMETERS FOR CATALYTIC FLAME STABILIZATION

FIGURE 8

FLAMEHOLDER DESIGN

The overall purpose of a flameholding device is to slow down the velocity of the exhaust jet stream entering the afterburner in the vicinity of the fuel supply so that a flame front may be established and remain stable. If the velocity of the stream is faster than the rate of flame propagation the stream will literally blow the flame out of the exit. If slower the flame front will try to propagate back upstream and blow itself out for lack of fuel. The problem then is to provide a device that will stabilize the overall combustion processes by providing a continuous pilot for the flame propagation wave.

The disadvantages of the bluff body flameholders are as follows:

1. High stream blockage with attendant high pressure drop across the flameholder.
2. No means of modulating the flameholder for various power settings.
3. Overheating of the flameholder may become a problem with the trend toward higher turbine inlet temperature.

The use of ceramic substate materials with a catalytic coating for flameholding should alleviate some of the above problems. Recent automotive emission requirements have provided the incentive for the production of various ceramic substate materials that are durable at high temperatures. High temperature ceramic materials that are currently available are:

Cordierite ($2\text{MgO} - 2\text{Al}_2\text{O}_3 - 5\text{SiO}_2$)

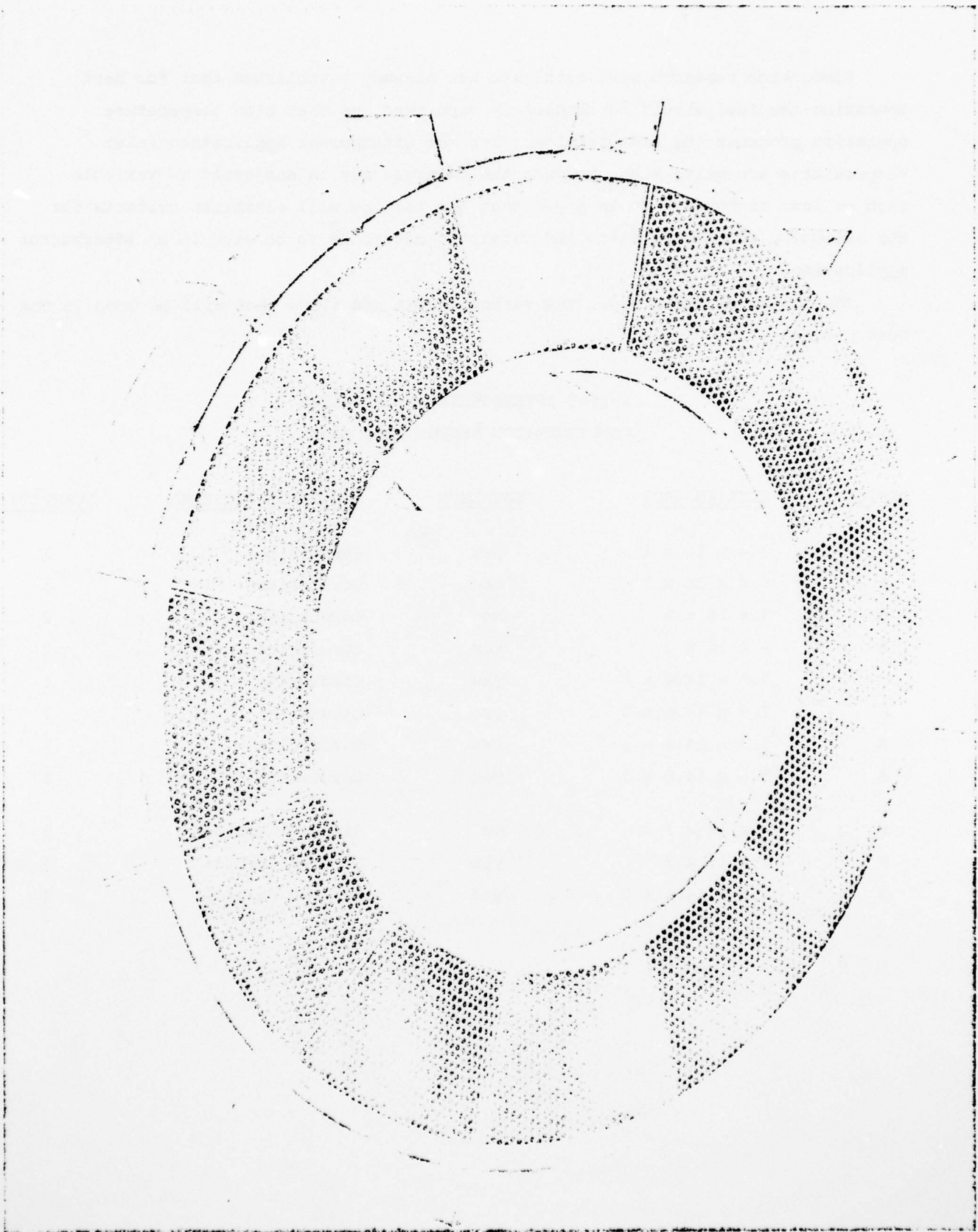
Silicon Carbide (SiC)

Alumina (Al_2O_3)

Zirconia (ZrO_2)

The materials selected for the J-85 flameholder study were Cordierite and silicon carbide. Of the four listed above cordierite has the highest strength and shock resistance with silicon carbide having the lowest strength and shock resistance. Thus these two materials will provide data at opposite extremes.

Catalyst surface coating materials that have been effective in promoting combustion and minimizing emissions include: platinum, palladium, iridium, cobalt oxide, and chromia. Platinum and palladium, materials used in automotive catalytic converters, have been selected as the coating materials for the flameholder tests. Several different sizes (inside and outside diameter) and thicknesses will be tested. The catalytic flameholder construction to be used is shown in Figure 9.



J-85- EXPERIMENTAL CATALYTIC FLAMEHOLDER

FIGURE 9

Combustion research with catalysts has already established that for best operation the fuel should be completely vaporized and that high temperature operation produces the best results. For the afterburner application inlet temperatures are quite high, however the catalyst may be subjected to variable rich or lean mixtures. It is hoped that the testing will establish criteria for the selection of both substrate and catalytic materials to be used in an afterburner application.

The following table shows the various types and sizes that will be used in the test program.

J-85-5 AFTERBURNER FLAME HOLDER
CONFIGURATION DESIGN AND SIZES

<u>DESIGN</u>	<u>SIZE (INCHES)</u>	<u>CATALYST</u>	<u>SUBSTRATE MATERIAL</u>	<u>QUANTITY</u>
A	8.4 x 14 x 1	yes	Cordierite	1
A	8.4 x 14 x 1	no	Cordierite	1
B	8 x 14 x 2	no	Cordierite	2
B	8 x 14 x 2	yes	Cordierite	2
C	7.6 x 14.6 x 1	yes	Cordierite	1
D	7.6 x 14.6 x 2	yes	Cordierite	2
E	5.9 x 14.8 x 1	yes	Cordierite	1
F	5.9 x 14.8 x 2	yes	Cordierite	2
B	8 x 14 x 2	no	Silicon Carbide	3
B	8 x 14 x 2	yes	Silicon Carbide	3
D	7.6 x 14.6 x 2	yes	Silicon Carbide	3

COMBUSTION OF A HYDROCARBON FUEL WITH AIR

Dry air is a mixture of gases with the following volumetric analysis in per cent:

O ₂	-	20.99
N ₂	-	78.03
A	-	0.94
CO ₂	-	0.03
H ₂	-	0.01

Argon listed in the above includes traces of neon, helium and krypton.

For most calculations dry air can be considered as 21 percent by volume oxygen and 79 percent nitrogen. Moisture content varies widely depending upon local conditions. The moisture content will add an additional amount of essentially inert material.

These calculations will consider the following:

$$.21 \text{ moles of O}_2 + .79 \text{ moles N}_2 = 1 \text{ mole air}$$

M_a - molecular weight of air

M_f - molecular weight of fuel

M_n - molecular weight of nitrogen

$$M_a = 28.964$$

$$M_n = 28.161 \quad (\text{This is weighted to include all of the inert gases})$$

Humidity will be considered negligible.

Fuel will be considered as C₁ H_{HCR}.

where: HCR - hydrogen/carbon ratio in the fuel.

$$\text{and } M_f = 12 + \text{HCR}$$

The combining equation, with most of the gaseous constituents expected in the exhaust from a turbojet engine, on a mole of air basis is as follows:

$$\text{FAR} \frac{M_a}{M_f} C \text{ HCR} + .21\text{O}_2 + .79\text{N}_2 = a \text{ CO}_2 + b \text{ H}_2\text{O} + c \text{ CO} + d \text{ H}_2 + e \text{ CH}_4 + f \text{ O}_2 + .79\text{N}_2$$

where: FAR - fuel air ratio by mass

HCR - hydrogen - carbon ratio in the fuel

a, b, c, d, e, f - mole volumes of CO₂, H₂O, CO, H₂, CH₄ and O₂ respectfully.

Unburned hydrocarbons are considered as CH₄.

Balancing the atoms of each constituent:

$$\text{C: } (\text{FAR}) \frac{M_a}{M_f} = a + c + e \quad (1)$$

$$\text{H: } (\text{FAR}) \frac{M_a}{M_f} (\text{HCR}) = 2b + 2d + 4e \quad (2)$$

$$\text{O: } .21(2) = 2a + b + c + 2f \quad (3)$$

$$\text{MT} = a + b + c + d + e + f .79 \quad (4)$$

where: MT - Total moles of exhaust per mole of air

The exhaust constituents that are measured are as follows:

CO₂ - in percent by volume on a dry basis

CO - ppm by volume on a dry basis

H₂ - ppm by volume on a wet basis

CH₄ - ppm by volume on a wet basis

H₂O, and O₂ are not measured. Considering that generally for any particular fuel HCR is known, examination of the above four equations reveals four unknowns namely FAR, b, f, and MT.

To solve the above four equations rewrite in terms of the constituents as measured.

$$a = \frac{\text{CO}_2}{100} (\text{MT} - b)$$

$$c = \frac{\text{CO}}{10^6} (\text{MT} - b)$$

$$d = \frac{H_2}{10^6} (MT)$$

$$e = \frac{CH_4}{10^6} (MT)$$

N. B. Only CO₂ and CO taken on a dry basis.

Rewriting the atom balance in terms of the constituents as measured.

$$C: (FAR) \frac{M_a}{M_f} = \frac{CO_2}{100} (MT - b) + \frac{CO}{10^6} (MT - b) + \frac{CH_4}{10^6} (MT) \quad (5)$$

$$H: (FAR) \frac{M_a}{M_f} (HCR) = 2b + \frac{2H_2}{10^6} (MT) + \frac{4CH_4}{10^6} (MT) \quad (6)$$

$$O: .21(2) = \frac{2CO_2}{100} (MT - b) + b + \frac{CO}{10^6} (MT - b) + 2f \quad (7)$$

$$MT = \frac{CO_2}{100} (MT - b) + b + \frac{CO}{10^6} (MT - b) + \frac{H_2}{10^6} (MT) + \frac{CH_4}{10^6} MT + f + .79 \quad (8)$$

Express $b = f (MT, HCR, FAR)$ from the hydrogen balance equation (6)

$$2b = (FAR) (HCR) \frac{M_a}{M_f} - \frac{2H_2}{10^6} (MT) - \frac{4CH_4}{10^6} (MT)$$

$$b = \frac{(FAR) (HCR)}{2} \frac{M_a}{M_f} - \frac{H_2}{10^6} (MT) - \frac{2CH_4}{10^6} (MT)$$

Express $f = \text{function } (FAR, MT)$ from the oxygen balance equation (7)

substitute the above value for b

$$.42 = \frac{2CO_2}{100} (MT) - \frac{2CO_2}{100} \left[\frac{(FAR) (HCR)}{2} \frac{M_a}{M_f} - \frac{H_2}{10^6} (MT) - \frac{2CH_4}{10^6} (MT) \right] + \frac{(FAR) (HCR)}{2}$$

$$\frac{M_a}{M_f} - \frac{H_2}{10^6} (MT) - \frac{2CH_4}{10^6} (MT) + \frac{CO}{10^6} (MT) - \frac{CO}{10^6} \left[\frac{(FAR) (HCR)}{2} \frac{M_a}{M_f} - \frac{H_2}{10^6} (MT) - \frac{2CH_4}{10^6} (MT) \right] + 2f$$

solving for f , number of moles of O₂

$$f = .21 + (A) \left(\frac{M_a}{M_f} \right) \frac{(FAR) (HCR)}{2} + (B) \frac{MT}{2}$$

$$\text{where: } A = \frac{CO_2}{100} + \frac{CO}{2 \times 10^6} - \frac{1}{2}$$

$$B = \frac{H_2}{10^6} + \frac{(2)CH_4}{10^6} - \frac{(2)CO_2}{100} - \frac{2(CO_2)H_2}{10^8} - \frac{(4)(CO_2)(CH_4)}{10^{12}} - \frac{CO}{10^6} - \frac{(CO)(H_2)}{10^{12}} - \frac{(2)(CO)(CH_4)}{10^{12}}$$

Note: A, B, C, D etc. will refer to groups of quantities obtained from test data

Express MT = f(b, MT) using the carbon balance equation (5)

$$(FAR) \frac{M_a}{M_f} = \left[\frac{CO_2}{100} + \frac{CO}{10^6} + \frac{CH_4}{10^6} \right] MT - \left[\frac{CO_2}{100} + \frac{CO}{10^6} \right] b$$

substitute b obtained from the hydrogen balance

$$(FAR) \frac{M_a}{M_f} = \left[\frac{CO_2}{100} + \frac{CO}{10^6} + \frac{CH_4}{10^6} \right] MT - \left[\frac{CO_2}{100} \times \frac{M_a(HCR)}{M_f} \right] FAR + \left[\frac{CO_2}{100} \times \frac{H_2}{10^6} \right] MT + \left[\frac{CO_2}{100} \times \frac{2CH_4}{10^6} \right] MT - \left[\frac{CO}{10^6} \times \frac{M_a(HCR)}{M_f} \right] FAR + \left[\frac{CO}{10^6} \times \frac{H_2}{10^6} \right] MT + \left[\frac{CO}{10^6} \times \frac{2CH_4}{10^6} \right] MT$$

Rearranging and solving for FAR

$$FAR = \frac{(C) M_f (MT)}{(D) M_a} \quad \text{Fuel air ratio}$$

$$\text{where: } C = \frac{CO_2}{100} + \frac{CO}{10^6} + \frac{CH_4}{10^6} + \frac{(CO_2)(H_2)}{10^8} + \frac{(2)(CO_2)(CH_4)}{10^8} + \frac{(CO)(H_2)}{10^{12}} +$$

$$\frac{(2)(CO)(CH_4)}{10^{12}}$$

$$D = 1 + \frac{(HCR)CO_2}{200} + \frac{(HCR)CO}{2 \times 10^6}$$

Now solve the above for MT (Total moles of exhaust per mole of air)

$$MT = \frac{(FAR)(D)(M_a)}{(C)(M_f)} \quad (9)$$

Substitute MT, b, and f in equation (8) and solve for FAR; first rearrange factoring out MT and b

$$\left[1 - \frac{CO_2}{100} - \frac{CO}{10^6} - \frac{H_2}{10^6} - \frac{CH_4}{10^6} \right] MT = \left[1 - \frac{CO_2}{100} - \frac{CO}{10^6} \right] b + f + .79$$

Now substitute MT, b, and f in the above

$$\begin{aligned}
 & \left[1 - \frac{\text{CO}_2}{100} - \frac{\text{CO}}{10^6} - \frac{\text{H}_2}{10^6} - \frac{\text{CH}_4}{10^6} \right] \frac{(\text{FAR}) (\text{D}) (\text{M}_a)}{(\text{C}) (\text{M}_f)} = \left[1 - \frac{\text{CO}_2}{100} - \frac{\text{CO}}{10^6} \right] \left[\frac{(\text{FAR}) (\text{HCR})}{2} \times \frac{\text{M}_a}{\text{M}_f} \right. \\
 & \left. - \frac{\text{H}_2}{10^6} \times \frac{(\text{D}) (\text{FAR}) \text{M}_a}{(\text{C}) (\text{M}_f)} - \frac{2 (\text{CH}_4)}{10^6} \times \frac{(\text{D}) (\text{FAR}) (\text{M}_a)}{(\text{C}) (\text{M}_f)} \right] + .21 \\
 & + A \left(\frac{\text{M}_a}{\text{M}_f} \right) \frac{(\text{FAR}) (\text{HCR})}{2} + B \frac{(\text{D}) (\text{FAR}) (\text{M}_a)}{(2) (\text{C}) (\text{M}_f)} + .79
 \end{aligned}$$

Grouping terms and solving for FAR (fuel air ratio)

$$\begin{aligned}
 \text{FAR} = \left(\frac{\text{M}_f}{\text{M}_a} \right) & \frac{1}{\left[\left(1 - \frac{\text{CO}_2}{100} - \frac{\text{CO}}{10^6} - \frac{\text{H}_2}{10} - \frac{\text{CH}_4}{10^6} \right) \frac{\text{D}}{\text{C}} - \frac{(\text{B}) (\text{D})}{(2) (\text{C})} + \frac{(\text{E}) (\text{H}_2) (\text{D})}{(\text{C}) 10^6} + (\text{E}) \frac{2 \text{CH}_4}{10^6} \times \frac{\text{D}}{\text{C}} \right.} \\
 & \left. - \frac{\text{A} (\text{HCR})}{2} - \frac{\text{E} (\text{HCR})}{2} \right]
 \end{aligned}$$

$$\text{where: } E = 1 - \frac{\text{CO}_2}{100} - \frac{\text{CO}}{10^6}$$

Substitute A, B, C, D, and E, cancel terms where possible and reduce the above to the following. Note that terms A, B, and E no longer appear, however a new group of terms called F appears.

$$\text{FAR} = \left(\frac{\text{M}_f}{\text{M}_a} \right) \frac{\text{C}}{(\text{F}) (\text{D}) + \frac{(\text{CO}) (\text{HCR}) (\text{C})}{4 \times 10^6} - \frac{(\text{HCR}) (\text{C})}{4}} \quad (10)$$

$$\text{where: } F = 1 - \frac{\text{H}_2}{2 \times 10^6} - \frac{\text{CO}}{2 \times 10^6} - \frac{(\text{CO}) (\text{H}_2)}{2 \times 10^{12}}$$

and to repeat the expression for the total moles of exhaust per mole of air.

$$\text{MT} = \frac{\text{M}_a}{\text{M}_f} \times \frac{(\text{FAR}) (\text{D})}{\text{C}} \quad (9)$$

For convenience in use the values for quantities C, D, and F will be repeated

$$C = \frac{\text{CO}_2}{10^6} + \frac{\text{CO}}{10^6} + \frac{\text{CH}_4}{10^6} + \frac{(\text{CO}_2) (\text{H}_2)}{10^8} + \frac{(2) (\text{CO}_2) (\text{CH}_4)}{10^8} + \frac{(\text{CO}) (\text{H}_2)}{10^{12}} + \frac{(2) (\text{CO}) (\text{CH}_4)}{10^{12}}$$

$$D = 1 + \frac{(\text{HCR}) (\text{CO}_2)}{200} + \frac{(\text{HCR}) (\text{CO})}{2 \times 10^6}$$

$$F = 1 - \frac{\text{H}_2}{2 \times 10^6} - \frac{\text{CO}}{2 \times 10^6} - \frac{(\text{CO}) (\text{H}_2)}{2 \times 10^{12}}$$

For convenience in checking results while setting up the main program both FAR(10) and MT(9) were programed and put on file. These two programs are presented in the appendix. MT or total moles is referred to as program DELTA. The fuel air ratio is referred to as program FAR.

TEST PROGRAM

The objective of the test program is to obtain and compare the performance of the J-85 afterburner using a conventional bluff body flameholder and several different catalyst coated and uncoated substrate screens for flameholders. The desired performance results are combustion efficiency and flameholder pressure drop. The major independent variable will be fuel flow (engine power setting). There will be four different power settings used for each test configuration. These will be designated as military, minimum afterburner, mid-afterburner and maximum afterburner. The independent variable at each power setting will be the exhaust probe position. A traverse will be made across the diameter, approximately eight inches from the engine exhaust nozzle. There will be eleven probe stations at the center of area of equal concentric areas across the exhaust plane.

Since it is virtually impossible to accurately obtain exhaust temperature measurements a complex mathematical procedure based on thermodynamic theory is used to obtain the exhaust temperature. This procedure being developed using a computer program will be summarized as follows:

The program calculates the local adiabatic flame temperature assuming 100% combustion efficiency. Using this temperature the program calculates equilibrium constants then calculates the number of moles of each exhaust gas constituent at equilibrium conditions. The program then takes the equilibrium conditions combined with the actual emissions and calculates the local actual combustion efficiency. Using this efficiency the program again uses an iterative procedure, as used to get the adiabatic flame temperature, and calculates what approaches a true total exhaust temperature. With this value the static temperature, density, and velocity can be calculated. Now local data is reduced using a least squares polynomial curve fit, integrating the results, normalized using the measured mass flow to finally obtain the overall emission concentrations, the calculated overall mass flow, and the overall combustion efficiency. A check on the accuracy of the test sampling and data reduction can be made by comparing the measured mass flow with the mass flow calculated from the emission data.

One of the problems in obtaining data is locating the edge of the exhaust stream. Generally this is taken as the radial location where the impact pressure equals the ambient pressure. However due to local turbulence producing vortices at the edge of the jet stream the aforementioned method may not accurately locate the edge of the exhaust stream. Another problem is in calculating the overall combustion efficiency based on the exhaust emissions. This is weighted by multiplying the local efficiency by the local mass flow and integrating the result across the exit area then dividing by the measured mass flow. This can best be shown by mathematical equations as follows:

$$(\text{Mass Flow})_m \eta = \int_0^R \eta_l d(\text{Mass Flow})_l$$

where:

η - overall efficiency

η_l - local efficiency

subscript m - measured

subscript l - local

$$d(\text{Mass Flow})_l = \rho (\text{Vel}) dA$$

$$dA = 2\pi r dr$$

$$(\text{Mass Flow})_l = 2\pi \int_0^R \rho (\text{Vel}) r dr$$

and

$$(\text{Mass Flow})_m \eta = 2\pi \int_0^R (\eta_l \rho \text{Vel}) r dr$$

$(\eta_l \rho \text{Vel})$ may be expressed as a function of the radius. This function is obtained using a least squares polynomial curve fit. Thus the final expression for the overall combustion efficiency based on sampling exhaust across the exit area is:

$$\eta = \frac{2\pi \int_0^R (\eta_l \rho \text{Vel}) (r) r dr}{(\text{Mass Flow})_m}$$

where: R - the effective exhaust stream radius

ρ - density in lbm per cuft.

Vel - local velocity in ft per sec

r - radius in feet

Mass Flow - in pounds per second

It can easily be seen from the above expression that this may lead to a problem, even though data may be acceptable, if the measured mass flow was higher than the integrated mass flow. This would cause the overall combustion efficiency to come out over 100 percent. To get around this problem and also to more accurately locate the edge of the exhaust stream an iterative look was put into the calculation program that causes the integration to proceed to a radius that will cause the measured mass flow to equal the integrated mass flow and thus keep the calculated efficiencies on the proper side of 100 percent. (As far as we know this method is unique to our project other research studies have generally used some arbitrary method in establishing the edge of the exhaust stream)

The calculation for ρV_{el} plays an important part in obtaining the calculated results. Some detail showing how calculations concerning ρV_{el} are handled is given in the appendix.

DATA PROCESSING

Twenty-eight test parameters are measured and recorded on magnetic tape by means of the data processing equipment. Preparation for actual testing involved obtaining a calibration curve for each channel and then getting a curve fit equation so that the voltage signal recorded on tape could then be converted to the actual parameter value. As an example of this the conversion procedure for the probe position and the nozzle position indicator are shown in the appendix. Copies of the calibration data are also included.

A program was written to read the engine operating and emissions data, print the data values and finally punch cards for input into a subsequent calculation program. The test parameters and outline format for fixed data, output, data summary, and program output are shown in the appendix.

APPENDIX
CALCULATIONS AND DATA SHEETS

AD-A065 650

OHIO STATE UNIV RESEARCH FOUNDATION COLUMBUS
USAF-ASEE (1978) SUMMER FACULTY RESEARCH PROGRAM (WPAFB). VOLUM--ETC(U)
NOV 78 C D BAILEY

F/G 1/3

F44620-76-C-0052

UNCLASSIFIED

AFOSR-TR-79-0231

NL

3 OF 6

AD
A065650



```

100=
110=C
120=C
130=
140=
150=
160=
170=
180=
190=
200=
210=
220=
230=
H2/1
240=
250=C
260=
270=
280=
290= 10
300=
310=
320=

PROGRAM FAR (INPUT, OUTPUT, TAPES=INPUT, TAPES=OUTPUT)
FAR- FUEL AIR RATIO

111=0
PRINT 4
FORMAT(3X,/,*,* ENTER NUMBER OF PROBLEMS.1,/,)
READ(5,*) NNN
1 ALPHA=2.02
PRINT 5
FORMAT(3X,/,*,* ENTER CO,CO2,CH4,H2%,/,)
READ(5,*) CO,CO2,CH4,H2
111=111+1
A=1.+ALPHA*CO2/200.+ALPHA*CO/2.E6
B=CO2/100.+CO/1.E6+CH4/1.E6+CO2*2.*CH4/1.E6+CO*
2.E12+CO*2.*CH4/1.E12
FAL=B/(1.-H2/2.E6-CO/2.E6-CO*H2/2.E12)*A/(CO*ALPHA*2)/4.E6
&B/4.)*(12.+ALPHA)/28.984
PRINT 10,FAL
FORMAT(3X,3FAL=*,F6.4,/,)
IF(111.LT.NNN) GO TO 1
STOP
END

```

```

100=
110=C
120=C
130=C
140=C
150=C
160=C
170=
180=
190=3
200=
210=5

PROGRAM DELTA (INPUT,OUTPUT,TAPES=INPUT,TAPES=OUTPUT)

DELTA=TOTAL MOLES EXHAUST PER MOLE AIR
LET ALPHA=2.02, 1 MOLE OF AIR=.21 MOLES OXYGEN + .79 MOLES
OF NITROGEN, MOL WT AIR=28.964, FUEL/AIR RATIO (FAL)

ALPHA=2.02
PRINT 3
FORMAT (3X,XENTER FAL,CO,CO2,AND CH4X)
READ 5,FAL,CO,CO2,CH4
FORMAT (F5.4,1X,F4.0,1X,F4.2,1X,F4.1)

H2=CO/2.
DELT=(1.+ALPHA*28.964/4./((12.01+ALPHA)*FAL*(1.-CO/1.E6)))
1/(1.-.5*(CO+H2)/1.E6-(H2-2.*CH4)/1.E6*CO/2.E6)
DELT=(1.+ALPHA*CO2/200.+ALPHA*CO/2.E6)*(FAL)*(28.964/(12.+AL
PHA))
1/(CO2/100.+CO/1.E6+CH4/1.E6+CO2*H2/1.E6+CO2*2.*CH4/1.E6+CO*H2
/1.
2E12+CO*2.*CH4/1.E12)
PRINT 10,DELT,DELT
FORMAT(3X,*DELT = *.F10.4,/,2X,*DELT = *.F10.4)
STOP
END
220=
230=
240=
250=
260=
270=
280=
290=10
300=
310=

```


Density And Velocity Calculations

$$\rho = \frac{P}{RT} \quad \text{Perfect gas relationship}$$

ρ - density lbm/ft^3 T - Temperature (absolute)

R - gas Constant

P - Pressure

$$Vel = \sqrt{\gamma g RT}$$

Same Velocity at exit

γ - ratio of specific
heat at const. ~~Pressure~~

$$\rho Vel = \frac{P}{RT} \sqrt{\gamma g RT}$$

to Sp. ht. at const. Vol.

$$\rho Vel = P \sqrt{\frac{\gamma g RT}{(RT)^2}} = P \sqrt{\frac{\gamma g}{RT}} \times \frac{\sqrt{\gamma g}}{\sqrt{\gamma g}}$$

$$\rho Vel = \frac{P \gamma g}{\sqrt{\gamma g RT}}$$

Now solve with P in psi, $g = 32.174 \text{ ft/sec}^2$

T in $^{\circ}\text{Kelvin}$

$$R = \frac{1.98717 \text{ cal} \cdot \text{ft} \cdot \text{lb}}{(\text{g} \cdot \text{mole})^{\circ}\text{K} \cdot 0.324083 \text{ cal}} = 6.13167 \frac{\text{ft} \cdot \text{lb}}{\text{g} \cdot \text{mole}^{\circ}\text{K}}$$

$$\text{Mol. Wt Air} = 28.9654 \frac{\text{g} \cdot \text{m}}{\text{g} \cdot \text{mole}}$$

Universal Values for R

$$R = \frac{6.13167 \text{ ft}^2 \text{ #}_f}{\text{# mole } ^\circ K} \times \frac{453.59237 \text{ gm}}{\text{#}_m}$$

$$\text{and } R = 2781.2787 \times \frac{1}{M.W.} \frac{\text{ft}^2 \text{ lbf}}{\text{lbm } ^\circ K}$$

M.W. - Mol. Wt. in gm/gmole or lbm/lbmole.

$$\therefore c_{Vel} = \frac{144 P \gamma (32.174)}{\sqrt{\gamma (32.174) \left(\frac{2781.2787}{M.W.} \right) T}}$$

with T in $^\circ K$

check Units

$$c_{Vel} = \frac{\frac{\text{lbf}}{\text{ft}^2} \frac{\text{ft}}{\text{sec}^2}}{\sqrt{\frac{\text{ft}}{\text{sec}^2} \times \frac{\text{ft}^2 \text{ lbf}}{\text{lbm } ^\circ K} \times ^\circ K}} = \frac{\text{lbf}}{\text{sec ft}^2}$$

Universal Value for R with T in $^\circ R$

$$R = \frac{1545.32}{M.W.} \frac{\text{ft}^2 \text{ lbf}}{\text{lbm } ^\circ R}$$

To check $e(Vel)$ use Max Power GE.
Run # 4-3 ; Run on Computer 7/10/78

ex. Local Point: X coordinate - .77 inches

$$\text{Static } T = 2906.07^{\circ}R$$

$$P^* = 13.93 \text{ psia}$$

$$\gamma = 1.255$$

$$\text{Av. Mol. Wt.} = 28.7215$$

For T in $^{\circ}R$

$$e \text{ Vel} = \frac{144(13.93)(1.255)(32.174)}{\sqrt{1.255(32.174)\left(\frac{1545.32}{28.7215}\right)2906.07}} = 32.235$$

this checks computer value

For T in $^{\circ}K$

$$T_K = \frac{2906.07}{1.8} = 1619.5$$

$$e \text{ Vel} = \frac{144(13.93)(1.255)(32.174)}{\sqrt{1.255(32.174)\left(\frac{2781.2787}{28.7215}\right)1619.5}} = 32.236 \quad \text{checks}$$

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO

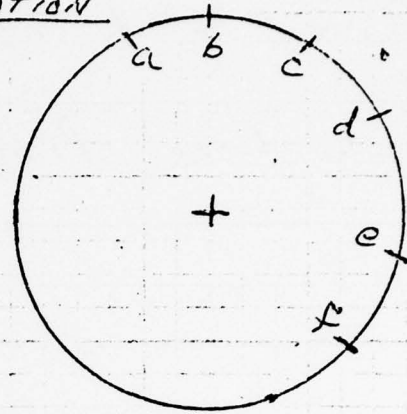
THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO

6/29/78

J-85-5 C-STAND

EXHAUST NOZZLE CALIBRATION

a, b, c, d, e, f. flat locations
not to scale. They are equidistant
from each other



To nullify the hysteresis Use Average
of the Opening And Closing Values.
N.E. hysteresis is low Max. Value = .15" @ 15%

Nozzle Opening %	Hysteresis Added to Opening	Average Dia. inches	Area in ²
2	+.02	10.77	91.1
10	-.10	11.33	100.82
20	-.04	11.99	112.91
30	-.075	12.58	124.29
40	-.018	13.19	136.64
50	0	13.80	149.57
60	+.025	14.40	162.86
70	-.04	15.0	176.71

$$\text{Area} = \frac{\pi D^2}{4}$$

Note: 80% is the
turnaround point
So 15 is not included in
the hysteresis table

SIGNATURE OF PROJECT OFFICER	70	15.54	181.67	D.R. Jensen	DATE	6/29/78
SIGNATURE OF WITNESS					DATE	
SIGNATURE OF WITNESS					DATE	

Rdg No.	Nozzle Opening No.	Inside Dia. Across Flats - inches						a+b+c +d+e +f	Average Dia. inches
		a	b	c	d	e	f		
1	2	10 $\frac{15}{16}$	10 $\frac{5}{8}$	10 $\frac{7}{8}$	10 $\frac{9}{16}$	10 $\frac{5}{8}$	10 $\frac{5}{8}$	64.56	10.76
2	10	11 $\frac{1}{2}$	11 $\frac{1}{4}$	11 $\frac{1}{2}$	11 $\frac{7}{32}$	11 $\frac{7}{16}$	11 $\frac{1}{4}$		11.53
3	20	12 $\frac{3}{4}$	11 $\frac{7}{8}$	12 $\frac{1}{8}$	11 $\frac{7}{8}$	12 $\frac{1}{8}$	11 $\frac{7}{8}$	72 $\frac{1}{8}$	12.01
4	30	12 $\frac{3}{4}$	12 $\frac{1}{2}$	12 $\frac{3}{4}$	12 $\frac{1}{2}$	12 $\frac{3}{4}$	12 $\frac{1}{2}$	75 $\frac{3}{4}$	12.55
5	40	13 $\frac{5}{8}$	13 $\frac{1}{4}$	13 $\frac{5}{8}$	13 $\frac{1}{2}$	13 $\frac{3}{8}$	13 $\frac{1}{2}$	77.19	13.17
6	50	13 $\frac{15}{16}$	13 $\frac{3}{4}$	13 $\frac{7}{8}$	13 $\frac{1}{4}$	13 $\frac{15}{16}$	13 $\frac{1}{4}$	82.82	13.8
7	60	14 $\frac{3}{4}$	14 $\frac{1}{4}$	14 $\frac{1}{2}$	14 $\frac{1}{8}$	14 $\frac{3}{8}$	14 $\frac{5}{8}$	86.31	14.385
8	70	15 $\frac{1}{4}$	14 $\frac{7}{8}$	15 $\frac{1}{8}$	14 $\frac{15}{16}$	15 $\frac{3}{8}$	14 $\frac{15}{16}$	90.25	15.02
9	80	15 $\frac{1}{2}$	15 $\frac{7}{8}$	15 $\frac{3}{8}$	15 $\frac{7}{16}$	15 $\frac{1}{2}$	15 $\frac{3}{8}$	93.25	15.59
10	70	15 $\frac{1}{8}$	14 $\frac{13}{16}$	15.0	14 $\frac{7}{8}$	15 $\frac{7}{16}$	14 $\frac{7}{8}$	77.87	14.98
11	60	14 $\frac{1}{2}$	14 $\frac{1}{4}$	14 $\frac{1}{2}$	15 $\frac{15}{16}$	14 $\frac{3}{8}$	14 $\frac{5}{16}$	86.43	14.41
12	50	13 $\frac{15}{16}$	13 $\frac{5}{8}$	13 $\frac{3}{8}$	13 $\frac{1}{16}$	14.0	13 $\frac{1}{16}$	82.82	13.8
13	40	13 $\frac{5}{8}$	13 $\frac{1}{4}$	13 $\frac{1}{4}$	13 $\frac{1}{4}$	13 $\frac{3}{8}$	13.0	79.06	13.18
14	30	12 $\frac{3}{4}$	12 $\frac{5}{16}$	12 $\frac{5}{8}$	12 $\frac{7}{8}$	12 $\frac{3}{4}$	12 $\frac{7}{16}$	75.26	12.59
15	20	12 $\frac{1}{8}$	11 $\frac{13}{16}$	12 $\frac{1}{16}$	11 $\frac{13}{16}$	12 $\frac{3}{16}$	11 $\frac{13}{16}$	71.81	11.97
16	10	11 $\frac{7}{8}$	11 $\frac{1}{8}$	11 $\frac{3}{8}$	11 $\frac{1}{4}$	11 $\frac{1}{2}$	11 $\frac{3}{8}$	67.97	11.37
17	2	10 $\frac{15}{16}$	10 $\frac{3}{8}$	10 $\frac{7}{8}$	10 $\frac{5}{8}$	11.0	10 $\frac{1}{16}$	64.7	10.76

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO

SIGNATURE OF PROJECT OFFICER

D. D. Jones

DATE

5/21/78

SIGNATURE OF WITNESS

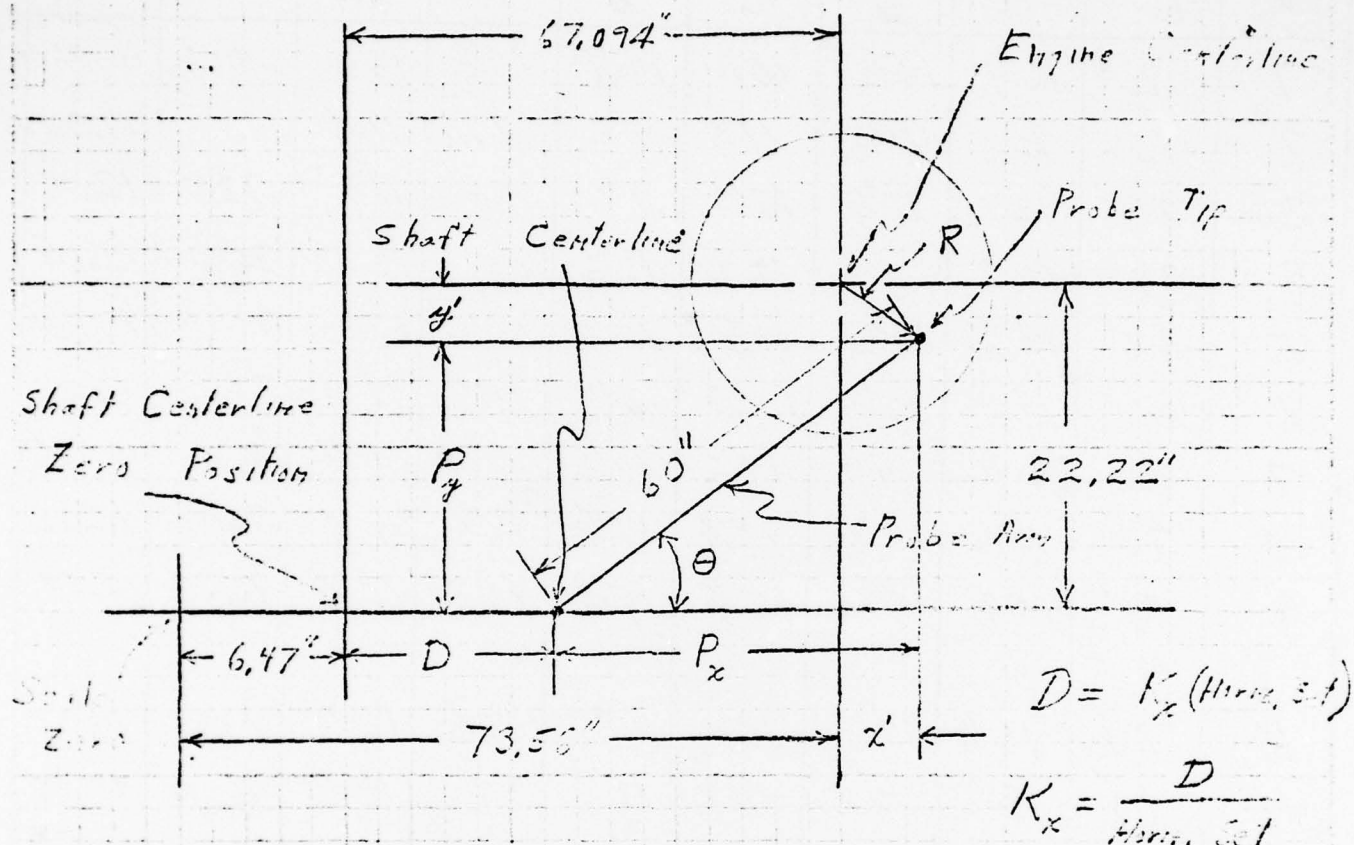
DATE

SIGNATURE OF WITNESS

DATE

6/2/72

Radial Probe Position Dimension With Respect to Engine Nozzle Centerline



D - Shaft Centerline X dist From Zero position

P_x - Probe dist from Shaft Centerline

P_y - Probe Y dist from Shaft Centerline

$$\theta = \cos^{-1} \frac{P_x}{L} = \cos^{-1} \left(\frac{P_x}{60} \right) \quad \left\{ \begin{array}{l} \theta = K_\theta (\text{angle set}) \\ K_\theta = \theta / \text{angle set} \end{array} \right.$$

$$x' = P_x + D - 67.094$$

$$y' = P_y - 22.22$$

$$R = \sqrt{(x')^2 + (y')^2}$$

SIGNATURE OF PROJECT OFFICER <i>D. F. [Signature]</i>	DATE 6/2/72
SIGNATURE OF WITNESS <i>[Signature]</i>	DATE
SIGNATURE OF WITNESS	DATE

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDQ

32

Rdg No.	Pencil mark Elev		D	P _x	Angle Scale Rdg.	θ degrees	Angle Conversion Constant K _{θ}	Horizontal Conversion Constant K _x
	Horizontal Set/Read	Vertical Set/Read						
Center line	3400/3400	1240/1200	11.3125	55.735	47.45	21.733	.017527	.0033272
1	1400/1400	1240/1200	4.6275	55.735	47.45	21.733	.017527	.0033482
2	2400/2400		8.0					.0033272
3	3400/3400		11.3125					.0033272
4	4400/4400		14.656					.0033331
5	5400/5400		17.989					.0033309
6	5400/5400		18.031					.0033331
7	4400/4400		14.688					.0033392
8	3400/3400		11.344					.0033336
9	2400/2400		8.031					.003346
10	1400/1400		4.688					.003349
							Σ	.0333725
							Avg K _x	.003339
1	2400/2400	1600/1600	11.3125	53.062	61.5	27.826	.017391	.0033272
2		1400/1400		54.625	53.6	24.437	.017455	
3		1240/1240		55.717	47.4	21.775	.017530	
4		1000/1000		57.125	33.0	17.809	.017819	
5		800/800		58.063	30.0	14.598	.018249	
6		300/300		58.125	30.0	14.362	.017753	
7		1000/1000		57.187	38.0	17.614	.017614	
8		1240/1240		55.750	47.5	21.645	.017496	
9		1400/1400		54.625	53.7	24.437	.017455	
10		1600/1600		53.062	61.5	27.826	.017391	
							Σ	.176373
							Avg K _{θ}	.017637
$\therefore \theta = 0.017637$ (Vert Rdg.)								
X DIST = D = 0.003339 (Horizontal Rdg.)								
THIS PAGE IS BEST QUALITY PRACTICABLE FROM COPY FURNISHED TO DDO								
SIGNATURE OF PROJECT OFFICER						DATE		
SIGNATURE OF WITNESS						DATE		
SIGNATURE OF WITNESS						DATE		

Notes For Check out Run To establish jet stream

Data Sheets

Blame + Bon 1 - Probe Total pressure
Pattern

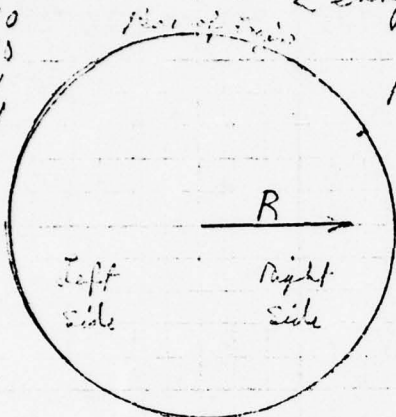
Lenny 2 - Control Room Panel rd.

Rich 3 - data processing info

R	Left Side	Right Side
5	4855	1890
5.5	5124	1740
6.0	5124	1595
6.5	5334	1441
7.0	5454	1291
7.5	5633	1141
8.0	5783	771

PPH

MIL -
 1500 A/B
 5500 A/B
 5000 A/B



$$P_{amb} = (Bar. P.) \times 4.91$$

$$XDIST = D = .003837 (11.31 \text{ ft})$$

$$XDIST = \frac{D}{11.31}$$

$$\Delta \text{ Honeys for } .5 \text{ miles} = 147.75$$

$$5' 150$$

R - Radius to probe

 $R = 11.31 - XDIST$ if $XDIST < 11.31$ Left Side

 $R = XDIST - 11.31$ if $XDIST > 11.31$ Right Side

For run at each of 4 different power settings
 Procedure at each power setting:

(a) Move probe to engine center. Read HPI and jet nozzle values

(b) Record control room panel readings data sheet # 2 and

data processing readings, data sheet # 3

(c) Run the probe out to $\frac{1}{2}$ " under the engine. Record total pressure

Move the probe out in $\frac{1}{2}$ " steps until probe total pressure
 is equal to ambient. Repeat on each side for mid, and mid A/B.

If necessary and practical run both sides for mid A/B and

SIGNATURE OF PROJECT OFFICER

D. R. Jenkins

DATE

7/18/78

SIGNATURE OF WITNESS

DATE

SIGNATURE OF WITNESS

DATE

THIS PAGE IS BEST QUALITY PRACTICABLE
 FROM COPY FURNISHED TO DDO

Run + Tag	Clock Time	Power Rating	Fuel Flow	NPI	Nozzle Radius	Probe Horizontal R.		Probe Total Pressure (psi. g)	(003339) X Horiz. Set X Dist	Radius to Probe
						Horizontal Set	Vertical Set			
Units			PPH	%	Inches			V _{1/2}		Inches
Eng. Center						3400	1240	24.23	11.31	0
0-0	2:45	1712	2500	7.5	2400	3000	3000	2.8718		
1-1	9:55	1712	2450	7.0	5.6	3400	1240	4.86	11.31	0
1-2	10:00		2420	6.0		4945		4.90	16.41	5.1
1-3	10:09		2420	6.5		5354		4.87	16.41	5.6
1-4	10:11		2420	7.0		5214		3.70	17.41	6.1
1-5	10:14		2420	7.0		5314		2.92	17.41	6.6
1-6	10:17		2450	6.5		5514		2.916	18.41	7.1
1-8	10:20		2420	7.0		1860		4.90		
1-9	10:22		2400	6.0		1710		4.30	5.71	5.6
1-10	10:25		2420	6.5		1560		3.14	5.21	6.1
1-11	10:27		2420	6.0		1411		2.88	4.71	6.6
1-12	10:29		2420	6.0		1261		2.267	4.21	7.1
2-0	10:33	1712 AB	1520	27.0	6.2	3400	1240	5.14	11.31	0
2-1	10:44		1510	26.5	1	5074		5.05	17.01	5.7
2-2	10:51		1440	26.5		5244		4.96	12.51	6.2
2-3	10:52		1470	26.0		5294		3.60	18.01	6.7
2-4	10:53		1430	26.0		5544		2.96	18.01	7.2
2-5	10:54		1470	26.0		5620		2.85	19.01	7.7
2-6	10:55		1460	26.0		1600		4.89	5.01	5.7
2-7	10:56		1490	26.0		1520		4.39	5.11	6.2
2-8	10:57		1490	26.5		1371		3.34	4.61	6.7
2-9	10:58		1500	26.0		1231		2.92	4.11	7.2
2-10	10:59		1460	26.0		1081		2.87	3.61	7.7
3-0	11:02	1712 AB	3460	50	6.9	3400	1240	5.15		0
3-1	11:07		3470	49		5304		5.13	17.71	6.4
3-2	11:10		3440	50		5454		4.94	18.71	6.9
3-3	11:11		3340	49		5603		3.63	18.71	7.4
3-4	11:12		3220	47		5753		2.92	19.71	7.9
3-5	11:15		3060	47		5903		2.86	19.71	8.4
3-6	11:19		3500	50		1470		4.95	4.21	5.4
3-7	11:20		3520	51		1321		4.43	4.71	6.0
3-8	11:21		3510	51		1171		3.47	3.71	
3-9	11:22		3520	50.5		1121		2.97	3.71	
3-10	11:23		3520	50.5		871		2.87	2.21	

SIGNATURE OF PROJECT OFFICER

[Signature]

DATE

July 18, 1970

SIGNATURE OF WITNESS

[Signature]

DATE

SIGNATURE OF WITNESS

DATE

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DOD

J-25.5

15 July 1972 Low Altitude

Start: 14:13
 Finish: 14:53
 End Time: 14:57

Bar. = $P_{\text{avg}} = 29.35 \text{ Hg}$

Run + Plp. No.	Clock Time	Power Setting	Fuel Flow	NPI	Nozzle Radius	Probe Radius Rds.		Probe Total Pressure (Absolute)	(2933.9) X Horiz. Set X DIST	Radius R to Probe
						Horizontal Set	Vertical Set			
Units			PPH	%	inches			Volts		inches
Engine Speed						3453	1240	16.23	11.31	0
4-0	14:13	MAX	4830			3003	3000	2.8735	2.3	
4-1	14:33		4830	64	7.4	3400	1240	4.645	11.31	0
4-2	14:40		4810	62		5454		4.52	12.21	6.9
4-3	14:42		4820	63		5603		4.25	13.71	7.4
4-4	14:44		4820	62.5		5753		3.34	17.21	7.9
4-5	14:45		4800	62		5903		2.95	17.71	8.4
4-6	14:46		4800	62		6053		2.87	20.21	8.9
4-7	14:47		4820	62		1321		4.31	4.41	6.9
4-8	14:48		4820	62.5		1171		3.97	3.11	7.4
4-9	14:49		4820	62		1021		3.21	3.41	7.9
4-10	14:49		4810	62		872		2.12	2.91	8.4
4-11	14:50	Y	4810	62.5	V	722	Y	2.37	2.41	8.9

Slope Line

H₂ - 5057

THIS PAGE IS BEST QUALITY PRACTICABLE
 FROM COPY FURNISHED TO DDC

SIGNATURE OF PROJECT OFFICER	DATE
SIGNATURE OF WITNESS	DATE
SIGNATURE OF WITNESS	DATE

ENGINE ON: 8:40

OFF: 8:49 - Oil Leak

ON: 9:09

24 JUL 78

RUN	TIME	POWER SETTING	NPI	NOZZLE RADIUS	TRANSVERSE PFS	Radius to Probe
1-0	9:23	MIL	6	5.65	-	
1-1	9:27		6.5		-3	5.1
1-2	9:31		6		0	0
1-3	9:33	Y	6		+3	5.1
TO A/B — 9:36						
2-1	9:38	MIN A/B	25	6.21	-5	7.55
2-2	9:42		25		-4	6.66
2-3	9:46		25		-3	5.63
2-4	9:48		24.5		-2	4.36
2-5	9:50		25		-1	2.52
2-6	9:52		25		0	0
2-7	9:54		25		1	2.52
2-8	9:57		25		2	4.36
2-9	10:00		25		3	5.63
2-10	10:04		25		4	6.66
2-11	10:06	Y	24.5	Y	5	7.55

COOLING 71% — 10:08

RUN	TIME	POWER SETTING	NPI	NOZZLE RADIUS	TRANSVERSE PFS	Radius to Probe
3-1	10:13	MID A/B	51	7.05	-5	3.29
3-2	10:16		51		-4	7.31
3-3	10:19		52		-3	6.13
3-4	10:21		51		-2	4.79
3-5	10:23		51		-1	2.76
3-6	10:25		52		0	0
3-7	10:27		53		1	2.76
3-8	10:30		53		2	4.79
3-9	10:32		53		3	6.18
3-10	10:36		53		4	7.31
3-11	10:38	Y	52	Y	5	8.29

COOLING 71% — 10:40

RUN	TIME	POWER SETTING	NPI	NOZZLE RADIUS	TRANSVERSE PFS	Radius to Probe
4-1	10:45	MAX A/B	64	7.40	-5	8.75
4-2	10:47		64		-4	7.72
4-3	10:49		65		-3	6.52
4-4	10:53		65		-2	5.05
4-5	10:55		64		-1	2.92
4-6	10:57		64		0	0
4-7	10:59		64		1	2.92
4-8	11:01		65		2	5.05
4-9	11:04		64		3	6.52
4-10	11:07		65		4	7.72
4-11	11:09	Y	64	Y	5	8.75

SHUT DOWN - 11:12

SIGNATURE OF PROJECT OFFICER	DATE
SIGNATURE OF WITNESS	24 JULY 1978
SIGNATURE OF WITNESS	DATE
SIGNATURE OF WITNESS	DATE

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO

PROBE POSITION		PROBE TOTAL PRESSURE		H ₂	H ₂
HORIZ	VERT	VOLTS	PSIA	(PPM)	(CH 9)
0	0	2.87	14.35	400	327
1872	1240	4.60	23.0	400	322
3399	1240	4.59	22.95	400	330
4927	1240	4.63	23.15	400	327

JET RADIUS = 6.8"
EFFECTIVE " = 7.22"

1138	1240	2.90	14.5	400	330
1405		3.55	17.75	470	350
1713		4.45	22.25	420	350
2093		4.65	23.25	430	346
2645		4.52	22.6	440	350
3399		4.59	22.95	435	351
4154		4.71	23.55	430	350
4705		4.60	23.0	430	348
5085		4.73	23.65	480	375
5394		3.28	16.4	430	340
5660	Y	2.86	14.3	400	316

JET RADIUS
7.55"

EFFECTIVE
RADIUS
7.96"

0	0	2.88	14.4		
916	1240	2.92	14.6	410	323
1210		4.08	20.4	415	340
1548		4.65	23.25	435	334
1965		4.70	23.5	440	363
2573		4.68	23.4	490	388
3399		4.67	23.35	450	364
4226		4.69	23.45	450	364
4834		4.68	23.4	470	379
5250		4.67	23.35	450	370
5588		3.70	18.5	470	373
5892	Y	2.86	14.3	400	320

JET RADIUS
8.29"

EFFECTIVE
RADIUS
8.74"

0	0	2.88	14.4		
779	1240	2.91	14.55	400	321
1087		3.83	19.15	415	332
1447		4.50	22.5	800	732
1887		4.56	22.8	1210	1134
2525		4.46	22.3	600	508
3399		4.42	22.1	510	432
4274		4.52	22.6	730	650
4912		4.55	22.75	3200	3160
5352		4.45	22.25	730	660
5711		3.58	17.9	425	340
6020	Y	2.88	14.4	400	315

JET RADIUS
8.75"

EFFECTIVE
RADIUS
9.22"

Note: Jet Radius obtained from previous plot of
Total pressure vs. distance. Effective radius is 10.00 inches.

SIGNATURE OF PROJECT OFFICER	DATE
SIGNATURE OF WITNESS	DATE
SIGNATURE OF WITNESS	DATE

7/27/73

J85-S TEST

To input data on tape

TIME	RUN	POWER	NOZZLE RADIUS (")	TRAVERSE PTS	RADIUS TO PROBE	HORIZ SET
	1-1	MIL	5.65 (6)	-5	6.85	1349
	1-2			-4	6.04	1590
9:32	1-3			-3	5.1	1872 ✓
	1-4			-2	3.95	2716
9:50	1-5			-1	2.28	2716
9:59	1-6			0	0	3799 ✓
10:04	1-7			1	2.28	4082
10:13	1-8			2	7.95	4582
10:16	1-9			3	5.1	4927 ✓
10:23	1-10			4	6.54	5209
10:33	1-11			5	6.85	5451

cool down 10:36

into A/B 10:40

10:41	2-1	Max A/B	7.40	-3	6.52	1447
10:46	2-2			0	0	3799
10:51	2-3			3	6.52	5352

cooling 10:57

shut down 11:04

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

SIGNATURE OF PROJECT OFFICER	DATE 7/27/73
SIGNATURE OF WITNESS <i>[Signature]</i>	DATE
SIGNATURE OF WITNESS	DATE

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO

CHANNEL NUMBERS

<u>OLD</u>	<u>NEW</u>	<u>PARAMETER</u>	<u>VARIABLE NAME</u>	<u>CALIBRATION CURVE</u>
1	1	E.G.T.	T5	$T5 = 45. + 900. * V / 4.2$
3	2	Inlet Temp	TO	$TO = 32. + 50. * V / 1.48$
5	3	Main Fuel	WFM	$WFM = 1000. * V$
6	4	A-B Fuel	WFAB	$WFAB = 1000. * V$
7	5	NO/NOX	NO	$NO = 0.1 * V * RNO$
8	6	CO	CO	$CO = 0.1 * V * RCO$
9	7	T.H.C.	THC	$THC = 0.1 * V * RTHC$
10	8	CO ₂	CO ₂	$CO_2 = 0.1 * V * RCO_2$
New	9	H ₂	H ₂	$H_2 = 0.1 * V * RH_2$
21	10	X Probe Pos	XDIST	$XDIST = 0.003333 * V + 0.337$
22	11	Y Probe Pos	ANGLE	$ANGLE = 0.17857 * V + 17.621$
2	12	NPI	A8	$A8 = V * 1.2334 + 98.585$
New	13	R.P.M.	NG	$NG = 16542.0 * V + 5.0$
11	14	Inlet Diff. #1	DP1(1)	$DP1(1) = V$
12	15	Inlet Diff. #2	DP1(2)	$DP1(2) = V$
13	16	Inlet Diff. #3	DP1(3)	$DP1(3) = V$
14	17	Inlet Diff. #4	DP1(4)	$DP1(4) = V$
15	18	Inlet Total #1	PT1(1)	$PT1(1) = 10. + V$
16	19	Inlet Total #2	PT1(2)	$PT1(2) = 10. + V$
17	20	Inlet Total #3	PT1(3)	$PT1(3) = 10. + V$
18	21	Inlet Total #4	PT1(4)	$PT1(4) = 10. + V$
4	22	Thrust	FG	$FG = 1000. * V$
20	23	Probe Tot. Press.	PT8	$PT8 = 5. * V$
19	24	New Press. (AMB)	PO	$PO = 10. + V$
New	25	New Press. (FLH)	PT5(1)	$PT5(1) = 10. * V$
New	26	New Press. (FLH)	PT5(2)	$PT5(2) = 10. * V$
New	27	New Press. (FLH)	PT5(3)	$PT5(3) = 10. * V$
New	28	New Press. (FLH)	PT5(4)	$PT5(4) = 10. * V$

NOTE: The ranges for emission instruments, RNO, RNO₂, etc. change with operating conditions. The table below indicates the instrument range for engine operating condition being run.

OPERATING CONDITION	RNO	RCO	RTHC	RCO ₂	RH ₂
MILITARY	50	2000	50	10	500
MIN A/B	50	20000	10000	10	5000
MID A/B	100	20000	5000	#	50000
MAX A/B	100	20000	2500	#	50000

$$\begin{aligned} \# \text{ CO}_2 &= .0007269 + .5457 * V + .20701 * V * * 2 \\ &- 1.036033 * V * * 3 + .0043749 * V * * 4 \\ &- .00014267 * V * * 5 \end{aligned}$$

$$R8 = 0.00778 * V * * 3 - 0.03751 * V * * 2 + 0.409 * V + 10.68$$

Fixed Data Format

<u>Date</u>	<u>Engine Condition Code</u>	<u>Configuration Code</u>			<u>Data Point</u>	
XXXXXX	XXX	XXX	XXX	XXX	XXX	XXX

Date Code

Day: 01 - 31
Month: 01 - 12
Year: 78 - 80

Engine Condition Code

Military Condition: 000
Min A/B Condition: 100 - 250
Mid A/B Condition: 250 - 400
Max A/B Condition: 410 - 600

Configuration Code (per cat. disk)

Design Type: 0 - Conventional Hardware
 1 - Design A
 2 - Design B
 3 - Design C
 4 - Design D
 5 - Design E
 6 - Design F
 7 - No Disk

Substrate Material: 1 - Corderite
 2 - Silicon Carbide

Catalyst Coating: 0 - No coating
 1 - Pt/Pd coating

Data Point Code

Data point: 000 - 100
Subpoint: 000 - 050¹¹

Output Format for Fixed Data

A/B Catalytic Flame Stabilization J85-5

June 21, 1978

Operating Condition - MAX A/B

Disk 1: Design C
Corderite
Pt/Pd Catalyst

Disk 2: Design D
Corderite
Pt/Pd Catalyst

Disk 3: No Disk

Data Point 001 Subpoint 001

<u>CHANNEL</u>	<u>QUANTITY</u>	<u>AVE VALUE</u>	<u>MAX</u>	<u>MIN</u>	<u>VOLTAGE AVE</u>	<u>MAX</u>	<u>MIN</u>
1	TS	1265.F	1268.F	1262.F	5.0	5.3	4.9
2	TO	60.F	60.F	60.F	3.2	3.2	3.2
3	WFM						
4	WFAB						
5	NO						
6	CO						
7	THC						
8	CO ₂						
9	H ₂						
10	XDIST						
11	ANGLE						
12	AB						
13	NG						
14	DP1 (1)						
15	DP1 (2)						
16	DP1 (3)						
17	DP1 (4)						
18	PT1 (1)						
19	PT1 (2)						
20	PT1 (3)						
21	TP1 (4)						
22	FG						
23	PT8						
24	PO						
25	PT5 (1)						
26	PT5 (2)						
27	PT5 (3)						
28	PT5 (4)						

The format on the preceding page shall be used for subpoints 2 thru n. Following subpoint n, an overall data summary shall be presented in the format shown on the succeeding page.

Following the data summary output a set of summary data cards shall be punched. The format shall be as shown below:

PUNCH 1000, DATAPT, T5, TO, WFM, WFAB, A8, NG, DP1(1), DP1(2), DP1(3), DP1(4)

1000 FORMAT (I3, F7.0, F7.1, F7.0, F7.0, ^{F7.0}F7.0, F7.4, F7.4, F7.4, F7.4)

PUNCH 1001, DATAPT, PT1(1), PT1(2), PT1(3), PT1(4), FG, PO, PT5(1), PT5(2), PT5(3), PT5(4)

1001 FORMAT (I3, F7.4, F7.4, F7.4, F7.4 F7.0, F7.4, F7.4, F7.4, F7.4, F7.4)

PUNCH 1002, SUBPT, RAD, CO, Co2, THC, H2, NO, NC2

FORMAT 1002 (I6, F7.4, F7.0, F7.2, F7.1, F7.1, F7.1, F7.1)

PUNCH 1003, SUBPT, RAD, CO, Co2, THC, H2, NO, NC2

FORMAT 1003 (I6, F7.4, F7.0, F7.2, F7.1, F7.1, F7.1, F7.1)

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO

✓ OVERALL DATA SUMMARY

✓ A/B CATALYTIC FLAME STABILIZATION J85-5

✓ JUNE 21, 1978

✓ OPERATING CONDITION - MAX A/B

Fixed Data

✓ Disk 1: Design C
Corderite
Pt/Pd Catalyst

Disk 2: Design D
Corderite
Pt/Pd Catalyst

Disk 3: No Disk

DATA POINT 001

✓ CHANNEL	QUANTITY	AVE VALUE	MAX	MIN	VOLTAGE AVE	MAX	MIN
1	T5						
2	TO						
3	WFM						
4	WFAB						
12	A8						
13	NG						
14	DP1(1)						
15	DP1(2)						
16	DP1(3)						
17	DP1(4)						
18	PT1(1)						
19	PT1(2)						
20	PT1(3)						
21	PT1(4)						
22	FG						
24	PO						
25	PT5(1)						
26	PT5(2)						
27	PT5(3)						
28	PT5(4)						

EMISSIONS DATA

Subpoint	Radial Position	5 PT8	6 CO	7 CO2	8 THC	9 H2	10 NO	11 NO2
1	6.77in	25.14psia	940ppm	9.60%	50ppm	25ppm	5ppm	10ppm
2	6.50in	26.00psia	920ppm	9.20%	50ppm	20ppm	5ppm	1ppm
↓	↓	↓	↓	↓	↓	↓	↓	↓

Changes in Channel Calibration Curves

Channel

Calibration Curve

10

$$X_{Dist} = 3.3370 * V$$

11

$$ANGLE = 17.6218 * V$$

12

$$RA8 = 0.00798V^3 - 0.03789V^2 + 0.409V$$

$$+ 10.688$$

13

$$NG = 12.5720 * 1/5.5$$

To Get Calibration Curve for Probe X and Y Position

Channel 10

See Calibration data
6/28/78

at center $, 003339 \times 3400 = 11.3526$

$K * V = 11.3526$

Condition	V	K
MIL	3.4020	3.3370
MIN	3.4021	3.3369
MID	3.4022	3.3368
MAX	3.4016	3.3374
		$\Sigma 13.3481$

$K = \frac{11.3526}{V}$

Av. $K = \frac{13.3481}{4} = \underline{\underline{3.3370}}$

Channel 11
at center

$\theta = 0.017637(1240) = 21.8699 \text{ degrees}$

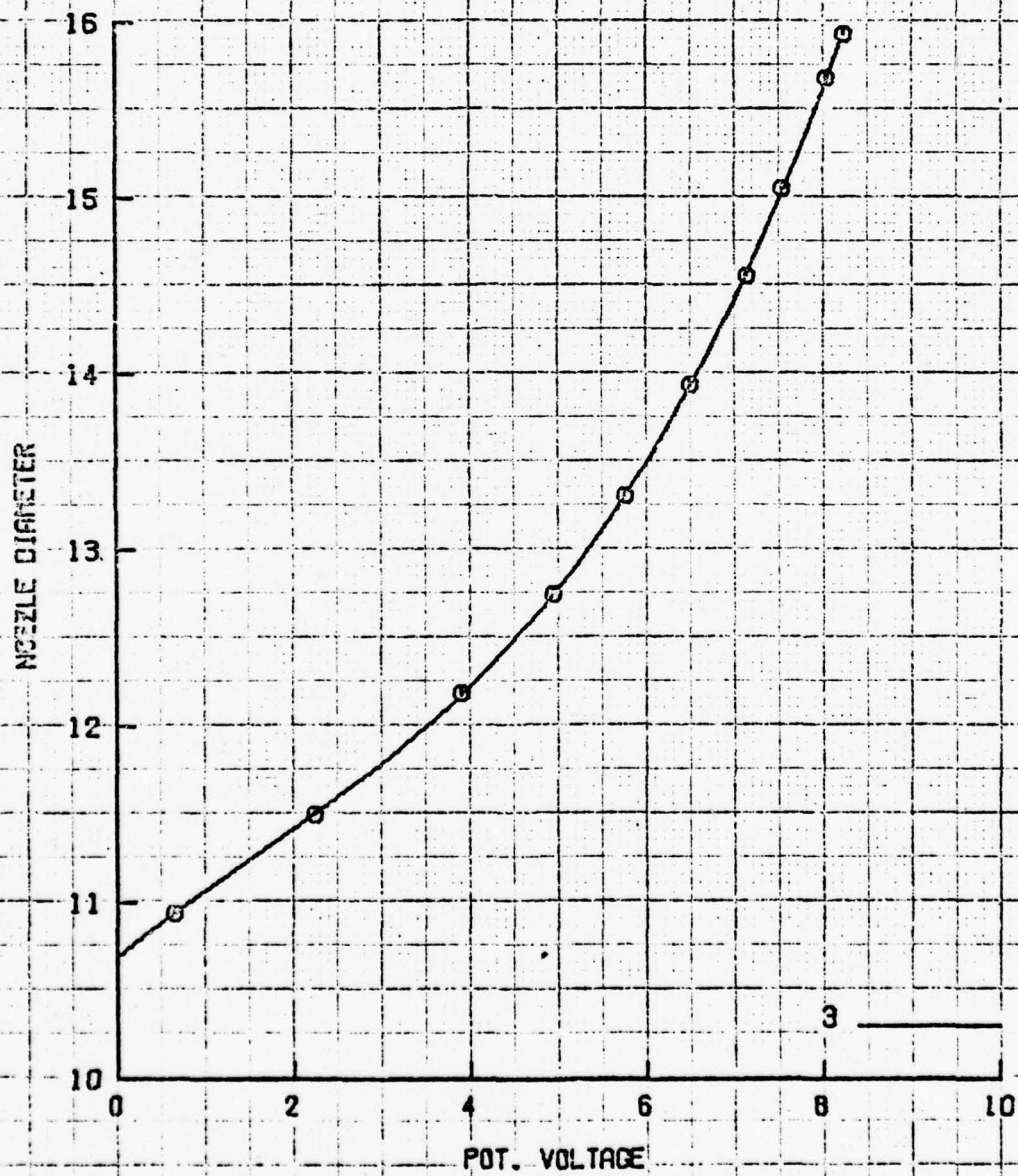
$\theta = h \times V \quad h = \frac{\theta}{V} = \frac{21.8699}{V}$

Condition	V	h
MIL	1.2416	17.6143
MIN	1.2411	17.6214
MID	1.2412	17.6200
MAX	1.2404	17.6313
		$\Sigma 70.4870$

$h = \frac{70.4870}{4} = 17.6218$

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

THIS PAGE IS BEST QUALITY PRACTICALLY
FROM COPY FURNISHED TO DDC



CALIBRATION FOR PROBE

JULY 1978

Channel 12

X

Y

1	.64346	10.94 00
2	2.24426	11.50 00
3	3.95940	12.19 00
4	4.93270	12.75 00
5	5.75340	13.31 00
6	6.50080	13.94 00
7	7.13710	14.56 00
8	7.54490	15.06 00
9	8.03500	15.69 00
10	8.23960	15.94 00

9.945827135E+00	6.6238474E-01		
1.100203440E+01	4.4931188E-03	7.099700143E-02	
1.560811517E+01	4.09015536E-01	-3.789023658E-02	7.97975421E-03
1.075475997E+01	3.77407535E-01	-2.292551605E-02	5.469221885E-03
1.075442688E+01	2.05020941E-01	5.631302155E-02	-1.61341961E-03
1.037011283E+01	1.26081031E+00	-7.005474021E-01	2.912430432E-01

EXLAX45 // // END OF LIST // //

Chiang 12

A/B CATALYTIC FLAME STABILIZATION J05-5

DATE

OPERATING CONDITION - MAX A/B

Disk 1: Design C	Disk 2: Design D	Disk 3: No Disk
Condition	Condition	
Pt/Pd catalyst	Pt/Pd catalyst	

Data Point 001

Channel	Quantity	Units	Ave. Vel.	Max	Min
---------	----------	-------	-----------	-----	-----

1

2

3

4

17

13

14

12

28

Page 2

A/B CATALYTIC FLAME STABILIZATION J85-

DATE

OPERATING CONDITION - MAX A/B

RAW EMISSIONS DATA

Radial

Position	PTA	CO	CO2	THC	H2	NO	NOX
----------	-----	----	-----	-----	----	----	-----

PROGRAM OUTPUT

LENNY

AIR FLOW

FUEL FLOW

FA

FAO

EMIN

EMISSIONS METHOD

EIEQ

TOTAL IDEAL TEMP., DEG. K

XEQ

EQUILIBRIUM VALUES OF PRODUCTS

EMIN

LOCAL

F/A

LOCAL MOL WT

TRUE
EFF

INTEG
EFF

TANAP

TOT EXCH TEMP

GAMMA

STATIC EXCH TEMP

MACH N

LEAST SQUARES CURVEFIT PROGRAM

ALL POINTS EQUALLY WEIGHTED

POLYNOMIAL COEFFICIENTS

RESULTS

X	Y	Y (MEAS)	RESIDUAL
---	---	----------	----------

RESIDUAL

INTEGRABLE LIMITS ARE:

AREA UNDER CURVE:

Additions

1) OUTPUT FIXED DATA

- need logic & format from tape program

2) OUTPUT FOR MIN AND MAX QUANTITIES

- need format from tape program

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO

TABLE
AVERAGE VALUES OF EMISSIONS MEASUREMENTS AT VARIOUS ENGINE POWER CONDITIONS

Test Conditions	No. of Data Points	Inlet Temp °K	Ambient Pressure Atms.	Mach No.	Gamma	Engine F/A	CO (%)	CO (PPM)	THC (PPM)	Local F/A	Comb. Eff.	Total Temp. °K	Static Temp. °K	Total Inlet Temp. °K	Static Inlet Temp. °K
450 RPM (1 S.D.)	10	286	0.980	0.11 +0.04	1.350 +0.005	0.0131 +0.0005	1.9 +0.2	2009 +436	557 +41	0.011 +0.001	0.924 +0.008	700 +39	698 +39	414 +39	412 +39
750 RPM (1 S.D.)	9	286	0.980	0.38 +0.02	1.356 +0.002	0.0094 +0.0002	1.77 +0.05	714 +103	165 +12	0.0089 +0.0002	0.972 +0.006	655 +10	639 +10	334 +11	339 +8
Military (1 S.D.)	13	289	0.980	1.08 +0.01	1.326 +0.001	0.0179 +0.0004	3.52 +0.01	468 +132	34 +11	0.0170 +0.0002	0.992 +0.002	970 +13	819 +19	691 +8	531 +11
1000 lbs A/B (1 S.D.)	16	292	0.977	1.070 +0.009	1.313 +0.002	0.0248 +0.0003	4.69 +0.03	2159 +308	1971 +297	0.024 +0.008	0.934 +0.008	1166 +43	1000 +45	878 +40	707 +40
1500 lbs A/B (1 S.D.)	3	282	0.978	1.060 +0.001	1.307 +0.001	0.0278 +0.0002	5.25 +0.01	1810 +52	2275 +90	0.0268 +0.0001	0.937 +0.002	1240 +1	1058 +2	959 +1	776 +2
2000 lbs A/B (1 S.D.)	14	292	0.977	1.052 +0.009	1.295 +0.006	0.0315 +0.0004	6.0 +0.2	2234 +332	1578 +344	0.0300 +0.0008	0.954 +0.01	1364 +25	1177 +27	1074 +22	866 +22
2500 lbs A/B (1 S.D.)	3	281	0.978	1.041 +0.001	1.292 +0.001	0.0344 +0.0001	6.54 +0.02	2367 +58	1112 +54	0.0323 +0.0001	0.965 +0.001	1439 +2	1242 +2	1158 +3	961 +3
3000 lbs A/B (1 S.D.)	12	292	0.977	1.034 +0.007	1.277 +0.005	0.0382 +0.0006	7.7 +0.3	1726 +270	336 +106	0.037 +0.001	0.985 +0.003	1610 +49	1405 +51	1320 +42	1115 +44
3500 lbs A/B (1 S.D.)	3	281	0.978	1.026 +0.001	1.274 +0.002	0.0407 +0.0002	7.99 +0.09	1450 +50	165 +4	0.0380 +0.0004	0.989 +0.001	1642 +12	1434 +11	1361 +12	1153 +11
4000 lbs A/B (1 S.D.)	14	292	0.977	1.023 +0.008	1.258 +0.007	0.0459 +0.004	9.1 +0.5	1110 +128	50 +18	0.043 +0.002	0.993 +0.002	1794 +65	1583 +66	1503 +60	1292 +61
4500 lbs A/B (1 S.D.)	3	282	0.978	1.014 +0.002	1.257 +0.001	0.0473 +0.0001	9.22 +0.04	1215 +25	44 +7	0.0432 +0.0002	0.993 +0.001	1806 +4	1554 +4	1524 +5	1313 +4
Max A/B (1 S.D.)	17	294	0.977	1.006 +0.006	1.237 +0.006	0.0518 +0.001	10.46 +0.4	2240 +258	32 +22	0.0439 +0.002	0.990 +0.001	1970 +47	1731 +48	1676 +43	1467 +44

INVESTIGATION OF CADMIUM DEPOSITION REACTIONS
ON

Cadmium and Nickel Electrodes

by

Cyclic Voltammetry

Yuen-Koh Kao

Technical Report

August 1978

AFAPL-POE-TM-78-10

W.U. 2303S402

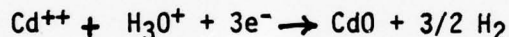
Aerospace Power Division
AF Aero Propulsion Laboratory
Wright-Patterson AFB OH 45433

Investigations of Cadmium Deposition Reactions on
Cadmium and Nickel Electrodes by Cyclic Voltammetry

Yuen-Koh Kao
University of Cincinnati, Cincinnati, Ohio

Abstract

Electrochemical impregnation is a promising method to produce cadmium electrodes for nickel cadmium batteries. Due to the complicated chemical and physical processes that are involved in the impregnation of the cadmium materials into the interstices of the porous nickel plaque material, wide-varying operating conditions has been reported in the literature. This work is an effort to understand the electrochemical reactions which occur during the impregnation process via cyclic voltammetry on nickel and cadmium electrodes in cadmium nitrate solutions. The reactions identified, for experiments on both nickel and cadmium electrodes, are the deposition of cadmium hydroxide, the reduction of cadmium hydroxide and the reduction of cadmium ions to cadmium. It is believed that the later reaction occurs only on the cadmium surface. The voltammetric data on the nickel electrode indicates that a passivating film is deposited on the nickel electrode. This species is believed to be cadmium oxide and could occur as the reduction reaction:



Nickel, by its nature is a good catalyst for the hydrogen evolution reaction, may catalyze the above reaction. It appears that the formation of this passive film could be deleterious to the Air Force electrochemical impregnation process for the manufacturing of cadmium electrode.

FOREWORD

This report presents the results of cyclic voltammetric studies of the reactions that occur during the electrochemical impregnation of cadmium electrodes. Dr. Yuen-Koh Kao, assistant Professor of Chemical Engineering at the University of Cincinnati was the principal investigator. This work was conducted during the summer of 1978 when he was a Faculty Fellow at the Ohio State University in the ASEE-USAF Summer Faculty Program (WPAFB). Acknowledgements are due to Dr. Joseph T. Maloy (AFAPL/Senior Investigator) for setting up the experimental apparatus and many helpful discussions, to Dr. John J. Lander AFAPL/POE-1 for his suggestions. Technician support was provided by Mr. John Leonard of the Aerospace Power Technical Support Branch (AFAPL/TFP).

Prepared by

YK Kao
Yuen-Koh Kao

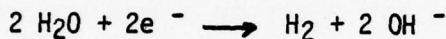
Reviewed by

R.A. Marsh
R.A. MARSH

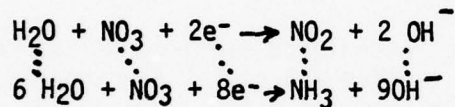
INTRODUCTION

Various methods of manufacturing cadmium electrodes exist in the patent literature. These methods can be classified into three categories (1): (a) impregnated nickel sinter type, (b) pressed powder type and (c) electrochemical impregnation. The last method is most attractive because it is simple in operation and can in theory achieve high loading of active material in a single impregnation cycle at lowest cost.

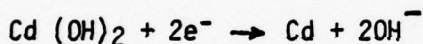
In the electrochemical disposition process, nickel plaque material serves as a cathode in an electrolysis cell. The cadmium or cadmium hydroxide is deposited onto the substrate from a cadmium nitrate solution. This deposition of $\text{Cd}(\text{OH})_2$ may be caused by the large amount of hydroxide ion near the electrode through either the direct liberation of H_2 at the electrode:



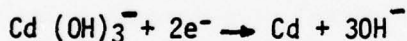
or the reduction of nitrate ion:



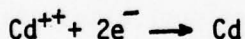
The deposited cadmium hydroxide may then be reduced to cadmium metal by the following reaction:



If pH near the electrode is high enough to favor the existence of cadmate ion, it can be reduced to cadmium by the following reaction:



Cadmium ions can also be reduced directly to cadmium metal:



Under suitable potentiostatic condition, one or more of the above reactions could occur during the impregnation process.

The success of electrochemical impregnation method has been achieved under wide-varying operating conditions. There are several variables in the electrochemical impregnation process. Beauchamp (2) in his patent emphasizes the fact that high impregnation bath temperature keeps the size of cadmium hydroxide crystals small and thereby obtaining an active electrode. The pH of the impregnation solution is another controlling factor. In Pickett's patent held by Air Force (3), the pH of the cadmium nitrate bath is maintained between 3-5 by adding proper amount of nitric acid. In Beauchamp's process, the pH is controlled by "buffering" the cadmium nitrate solution with sodium nitrite. The current density, based on apparent electrode area is another important controlling parameter, varies from 4-8 amps/m² in Beauchamp's

process (2) to as high as 1.2-1.6 amps/in² in the Pickett process (3). Bulen in his patented process (4) claims that the electrode impregnated by alternate current technique retains much of its capacity after cycling as compared to electrode fabricated by other methods.

The electrochemical impregnation of cadmium into the interstices of a porous substrate is very complicated involving both chemical and physical processes. The diversity of methods in the patent literature suggests that further improvement of the state-of-the-art of the said process is not possible without further understanding of the related electrochemical reactions. The understanding of these electrochemical reactions shall provide a baseline whereupon further optimization and refinement of the manufacturing process can be attempted.

The electrochemistry of the Cd/Cd(OH)₂ electrode in concentrated alkaline solution, typically KOH as in Ni/Cd battery, has been studied extensively. (5,6) There is, however, little information available about the electrochemical reaction in cadmium nitrate solutions.

OBJECTIVE

It is the objective of this study to investigate electrochemical reactions of Cd/Cd(OH)₂ deposition as occurring on cadmium and nickel electrode surfaces in Cd(NO₃)₂ solutions, especially the role that the nickel plays in the deposition process. Also investigated is the effect of pH to the electrodeposition process in the fabrication of cadmium electrodes.

EXPERIMENTAL

Cyclic voltametry experiments were conducted with a Princeton Applied Research Model 170 Electrochemical System. The nickel and cadmium micro-electrode were constructed in house at AFAPL. The nickel electrode was constructed by forcing a 0.05 inch diameter nickel wire through a Teflon sleeve and grinding the end flat. The cadmium electrode was constructed by encasing a cadmium strip in a lucite rod and again grinding the end flat. The square cadmium electrode had an area of 0.01 in². The reference electrode, also made in house at AFAPL, was an aqueous saturated calomel electrode fitted with a side arm salt bridge with a pin-hole junction at the end. The counter electrode was a strip of bare cadmium metal.

All experiments were conducted in a covered 400 ml beaker. The electrolyte was deaerated and held under helium atmosphere. The electrolyte was either 2M Cd(NO₃)₂ or 1M KNO₃. The pH of the solution is controlled by adding small amount of nitric acid. All experiments were conducted in a quiescent solution.

For each run, the electrode was polished with 4/0 energy paper immediately before use to assure a fresh metal surface. The electrode is then rinsed with distilled water before it was put into the cell.

RESULT AND DISCUSSION

To understand the role of nitrate ions in the deposition process, both nickel and cadmium microelectrode were subjected to cyclic voltammetry in the KNO_3 solution. The results are shown in Figures 1 and 2. There is a noted absence of any detectable faradic process in the range -0.4 to -0.8 vs. SCE.

Figure (1.a) is a voltammogram of nickel electrode in 2M $\text{Cd}(\text{NO}_3)_2$ and 1M KNO_3 solutions at pH = 2.6. The scan covers range between -0.4 to -0.8V vs. SCE, well into the cathodic background of nickel electrode. There is a significant faradic process occurs on the initial scan out to the cathodic direction. The potential where this process occurs varies with pH, from -0.624V vs. SCE at pH = 1.5 to -0.678V vs. SCE at pH = 6.0. This wave appears only at the initial scan and causes the passivation of the electrode surface. The surface species is believed to be cadmium oxide. More on this will be discussed below.

On the return scan, another faradic process occurs at -0.63V vs. SCE. This is believed to be the reduction of cadmium hydroxide to cadmium metal as was indicated by equation (3). This cadmium hydroxide is presumably deposited onto the electrode at negative potentials when large amounts of hydroxide ions are generated near the electrode. The cathodic background process could either be the reduction of nitrate ion and its product as indicated by equation (2) or the reduction of water evolving hydrogen as indicated by equation (1). When the background hydroxide ion generation reaction is reduced by sweeping less into the cathodic background as was done in the next experiment in which the sweep range was between -0.4 and -0.7V vs. SCE as shown in Figure (3.b). There is not hydroxide ions generated, so cadmium hydroxide is not precipitated and consequently, there is no current due to the reduction of cadmium hydroxide on the return scan. In addition, the cadmium oxide film cannot be stripped away and the electrode is passivated.

Figure (2.a) shows the multiple cyclic voltammogram of cadmium electrode in the same solution, 2M $\text{Cd}(\text{NO}_3)_2$ and 1M KNO_3 , at various pH = 2.6. The range of sweep was between -0.6 to -0.8V vs. SCE. The narrow scan range was chosen to prevent significant dissolution of cadmium electrode which would occur if the electrode were subjected to potential much more anodic than the rest potential, about -0.61V vs. SCE. There is a distinct difference between the voltammograms of the cadmium electrode and the nickel electrode shown in Figure 1. The faradic process that occurred on the initial scan of the nickel electrode that passivated the electrode is absent in the cadmium electrode.

There are three cathodic processes occurring on the cadmium at -0.63, -0.65 and -0.67 V vs. SCE. Two faradic processes occurred at -0.63 and -0.65V vs. SCE are believed to be the cathodic reduction of two different forms of cadmium hydroxides. The active and the inactive form of hydroxide standard reduction potentials are different by 17 mV (7). This process happens only on the return scan indicating that the reactant of this process is cadmium hydroxide deposited onto the electrode in the cathodic background of hydroxide ion generation.

The direct reduction of the cadmium ion is responsible for the current which occurred at -0.69V vs. SCE. This process does not appear on the nickel electrode experiments shown earlier in Figure 1. The reason behind this phenomenon is that this process can only happen on the cadmium surface. In the earlier experiments, on a nickel electrode, the cadmium surface that was created by the cadmium hydroxide reduction was stripped away in the anodic process. Therefore, the cadmium ion reduction process at 0.69V should occur if the cadmium generated in the nickel electrode was scanned between -0.6 to -0.8V vs. SCE except the first scan which starts at -0.4V vs SCE. The result shown in Figure (3) supports the above theory by the appearance of the reduction wave at -0.69V vs SCE when the anodic stripping of the cadmium ion was stopped. It is therefore concluded that the faradic process occurs at -0.69V vs SCE is due to the reduction of cadmium ion to cadmium metal and this process can occur only on a cadmium surface.

The faradic process that characterizes the initial scan in the cathodic direction which is responsible for the subsequent passivation of the nickel electrode is proposed to be the following process:



This process occurs only on the nickel electrodes and is believed to be catalyzed by nickel which is a good hydrogen scavenger. The cadmium oxide film formed cannot be removed electrochemically.

The nickel electrode was subjected to different sweep rates in electrolytes with pH value around 3.5 to study the controlling mechanism of the cadmium oxide formation. The results are shown in the following table:

Sweep Rate v (mV/sec)	Peak Current i_p (μA)	$i_p/v^{1/2}$
10	164	51.86
20	210	46.96
50	320	45.25
100	446	44.60
200	700	49.50

It is thereby concluded that the cadmium oxide formation reaction on the nickel electrode is controlled by the diffusion rate of either cadmium or hydronium ions.

The above cadmium oxide formation process depends on the pH of the electrolyte as shown in Figure (4). $E_{1/2p}$ values are -0.57V, -0.61V and -0.64 V vs SCE when the pH values of the electrolyte are 1.5, 3.3, and 6.0. The cadmium hydroxide reduction current does not depend on the pH of the electrolyte.

Figure (5) shows the pH dependence of the electrochemical reactions at the cadmium electrode. In general, reaction currents of all faradic processes increases as pH of the solution decreases.

CONCLUSION

The results of this study reveals the complicated sequence of reactions that takes place in the electrochemical impregnation process for the fabrication of cadmium electrode. A cathodic reduction process was identified in the initial cathodic scan of nickel electrodes in a cadmium nitrate solution. This process is argued to be the nickel catalyzed cadmium oxide formation, which is responsible for the passivation of the electrode. This process is pH dependent and becomes more cathodic in a more basic solution. There may be two reducible forms of cadmium hydroxide deposition at potential sufficiently negative for hydroxide formation as indicated by the two faradic processes that appear upon scan reversal. Direct reduction of cadmium ion occurs at a more cathodic potential than cadmium hydroxide and this reaction can happen only on a cadmium surface.

The above results has its implications in both the electrochemical impregnation process for making cadmium electrodes and the performance of a cadmium electrodes in an alkaline buffer. Lowering the pH reduces the reduction potential of the cadmium oxide formation on the nickel surface. This may be the reason that electrochemically impregnated cadmium electrodes formed under acidic condition and fails to retain its capacity upon cycling. The complicated sequence of electrochemical chemical reactions indicates the need of proper potentiostatic control of the impregnation process.

The nickel substrate probably plays an important role in the passivation of the cadmium electrode. Nickel is a good electrocatalyst for hydrogen evolution and may thereby promote the formation of the irreducible cadmium oxide film that passivates the electrode. The nickel surface when exposed to a cadmium negative electrode may cause the permanent loss of capacity by the formation of this passivating film during the changing cycle of the battery.

Since it is believed that hydrogen dissolves extensively in metallic nickel, this nickel electrode may actually catalyze reaction (6) by removing hydrogen gas as it is formed at the electrode surface. This conjecture is supported by the evidence (Figure 3) that no passivating film is formed on a nickel electrode that has a fresh cadmium surface deposited upon it by the reduction of cadmium hydroxide. It appears that the formation of this passive film is deleterious to the Air Force electrochemical impregnation process for making cadmium electrodes. Work should be initiated to make nickel plaque material behave like cadmium in the deposition process.

REFERENCES

1. David F. Pickett and Vincent Puglisi, "Electrochemical Impregnation of Sintered Nickel Structures with Cadmium Using Constant Current Step and Alternate Current Pulse Techniques", Technical Report AFAPL-TR-74-119 (1975).
2. R. L. Beauchamp, "Method for Producing a Cadmium Electrode for Nickel Cadmium Cells", U.S. Patent 3,573,101, (1971).
3. David F. Pickett, "Production of Cadmium Electrodes", U.S. Patent 3,873,368, (1975).
4. E. F. Bulen et.al., "Method of Making Electric Battery Electrodes", U.S. Patent 3,484,346, (1969).
5. Sidney Gross and Robert J. Glocking, "The Cadmium Electrode, Review of the Status of Research", The Boeing Company Report D180-19046-2, (1976).
6. P. McDermott et.al., "Secondary Aerospace Battery and Battery Materials, a Bibliography (1969-1974)", NASA SP-7044, (1976).
7. Marcel Pourbaix, "Atlas of Electrochemical Equilibria in Aqueous Solutions", Pergamon Press, New York, (1966).

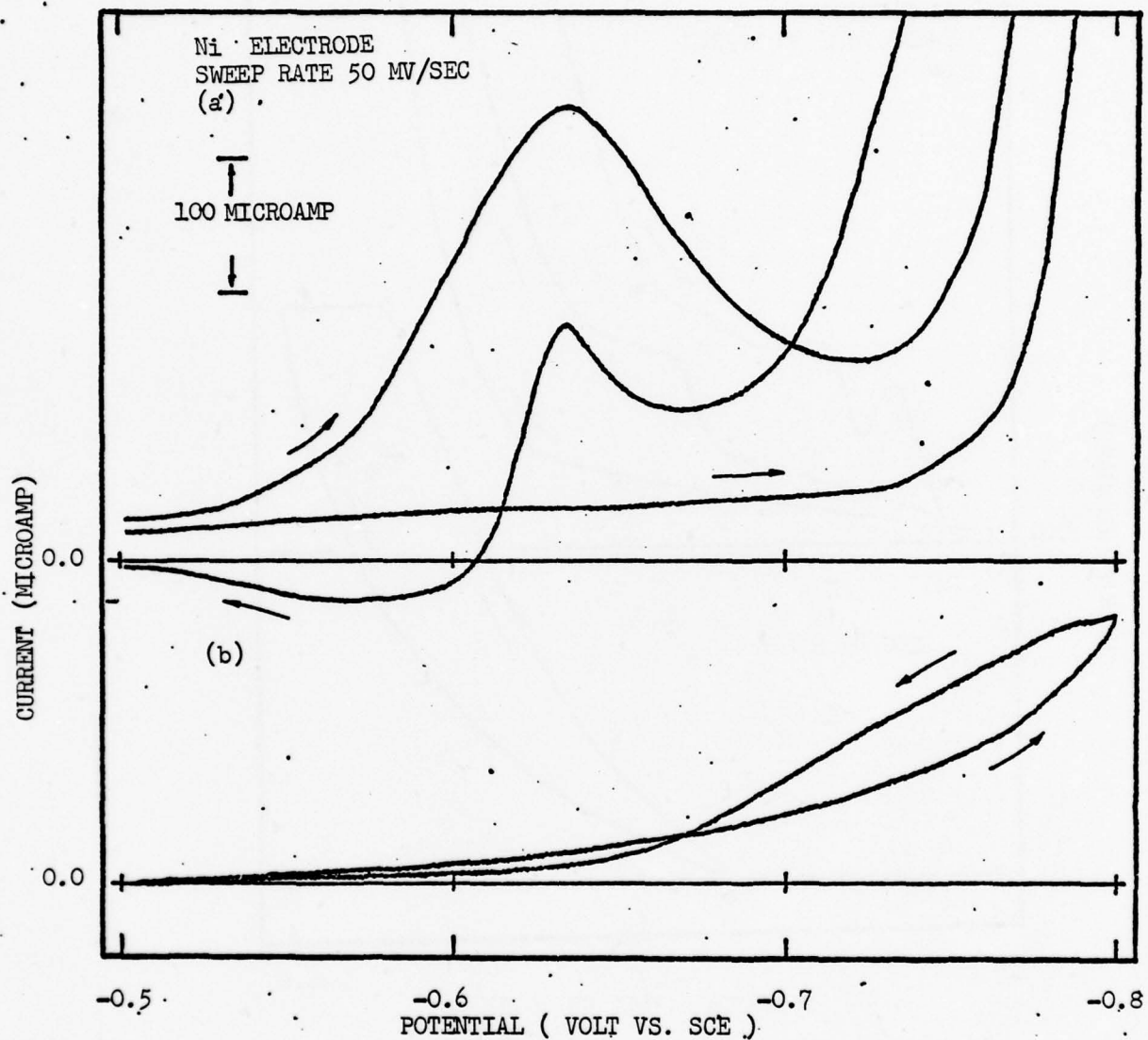


Figure 1: Multiple scan voltammetry of nickel electrode into the cathodic background in: (a) $2M \text{ Cd}(\text{NO}_3)_2$ $1M \text{ KNO}_3$ solution, and (b) $1M \text{ KNO}_3$ solution.

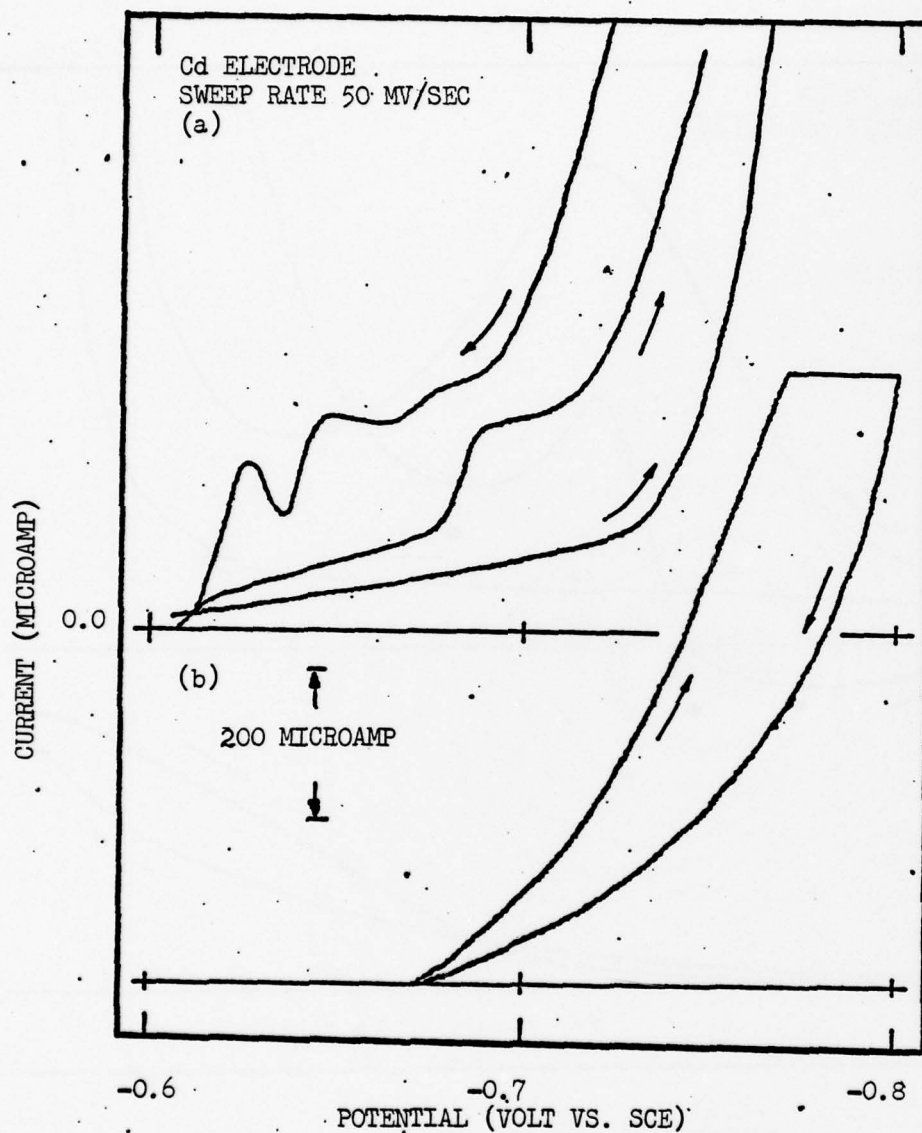


Figure 2: Multiple scan voltammetry of cadmium electrode into the cathodic background in: (a) 2M $\text{Cd}(\text{NO}_3)_2$ 1M KNO_3 solution, and (b) 1M KNO_3 solution.

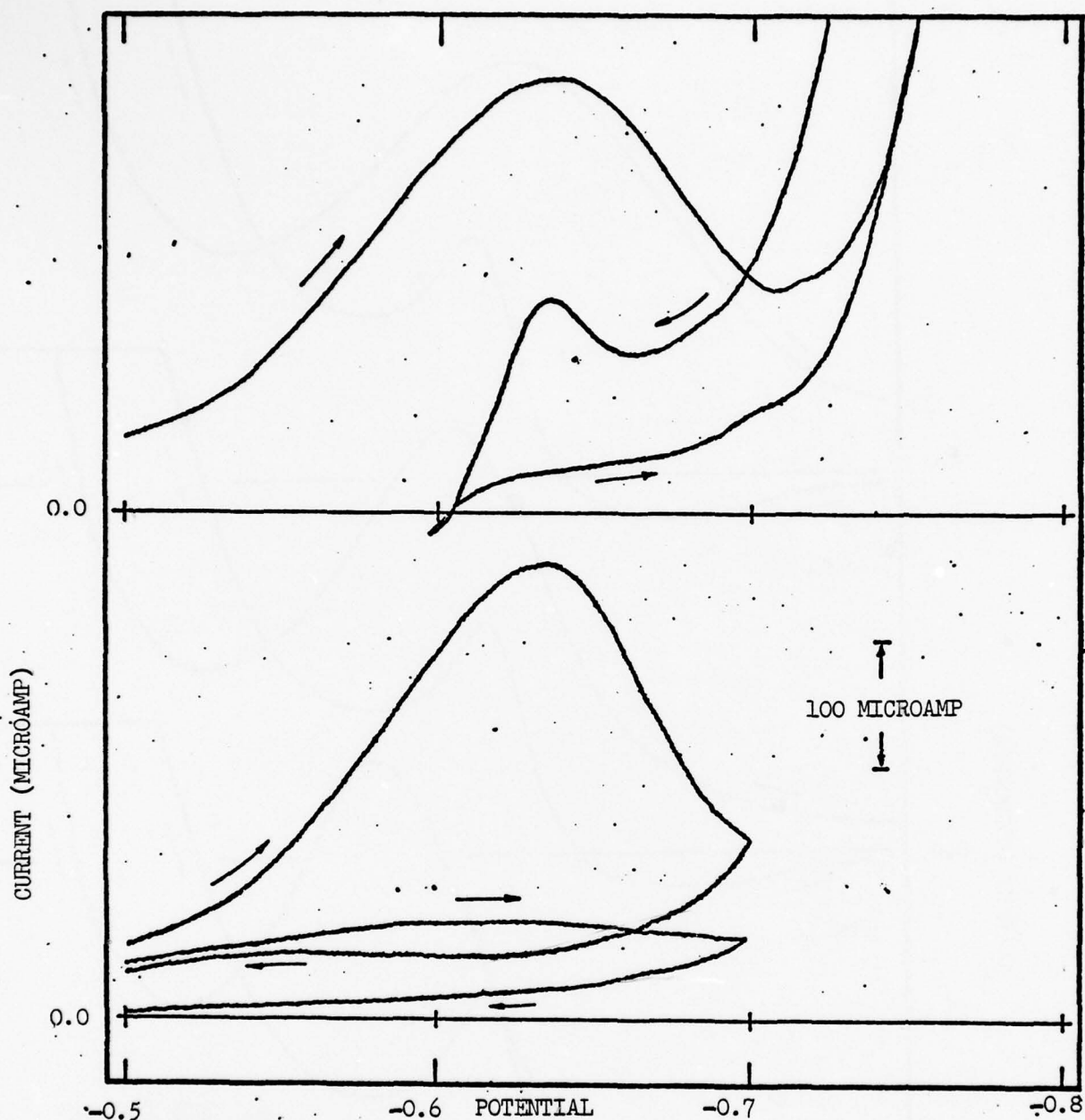


Figure 3: Multiple scan voltammetry of nickel electrode in 2M $\text{Cd}(\text{NO}_3)_2$ 1M KNO_3 solution.

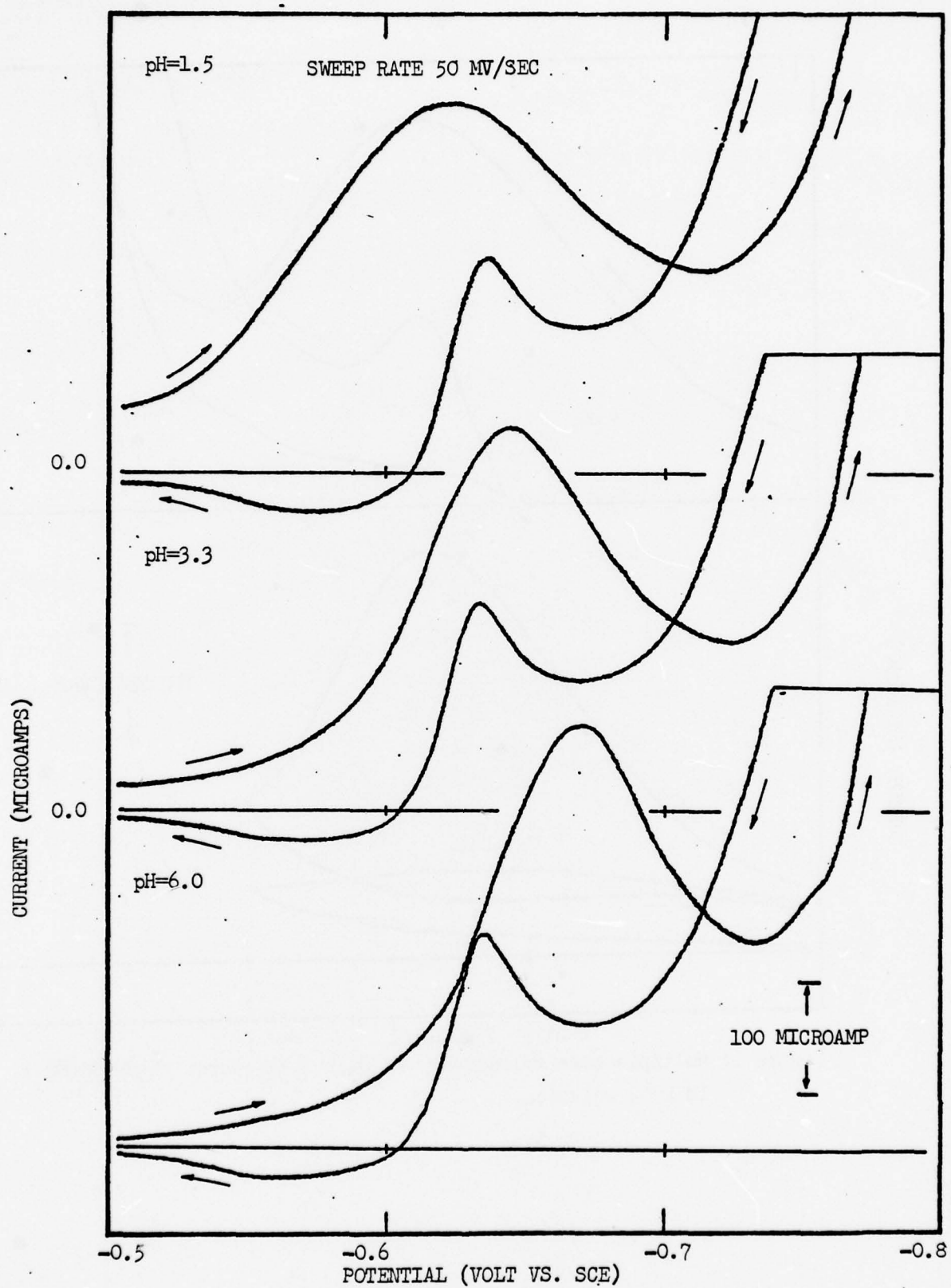


Figure 4: Multiple scan voltammetry of nickel electrode in 2M $\text{Cd}(\text{NO}_3)_2$ 1M KNO_3 solution.

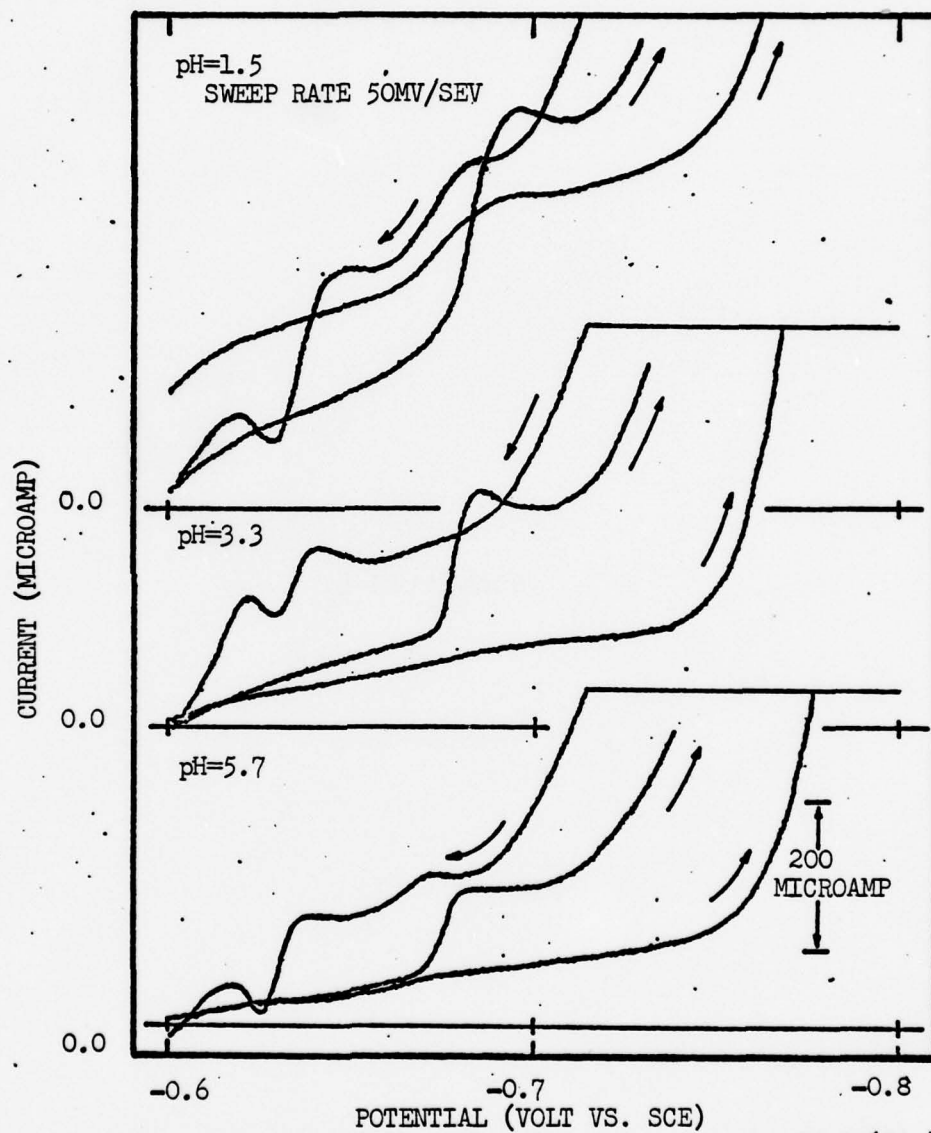


Figure 5: Multiple scan voltammetry of cadmium electrode in 2M $\text{Cd}(\text{NO}_3)_2$ 1M KNO_3 solution.

USAF-ASEE SUMMER FACULTY PROGRAM (WPAFB)
FINAL REPORT

Submitted by

Dr. Chris C. Lu
Department of Chemical Engineering
University of Dayton
August 1978

Dr. Chris C. Lu
Dept. of Chem. Engr.
Univ. of Dayton

USAF-ASEE Summer Faculty Program (WPAFB) - Final Report

Subject: Turbulent Flow Measurements for Sudden Expansion Cylindrical Tube by Using Laser Doppler Velocimeter (LDV)

Summary: The profiles of turbulent velocity and intensity for a sudden expansion cylindrical tube are measured at both center line and several cross sectional areas by using the LDV technique. The sudden expansion cylindrical tube is chosen in this experiment to simulate the ramjet combustion system. The scatter of velocity profiles is about 1% ~ 2% at the low turbulent intensity region and can be as high as 20% at the high turbulent intensity region, Figure 3. The turbulent intensity is still difficult to measure when it is high, because the perturbation of velocity is very random and fast at this region, Figure 4. In order to obtain good data for turbulent measurements, a computer is suggested to connect to a current existing LDV signal processing unit thus a large number of data can be analyzed by the computer.

I. Introduction

The technique of laser Doppler velocimeter (LDV) has been used quite often in fluid velocity measurement. Yeh and Cummins (1) introduced the technique of LDV in 1964 for the steady state laminar flow in a circular tube. In 1967, Goldstein and Kreid (2) demonstrated the LDV technique to measure laminar flow in a square duct. Since then, many investigators (3 - 10) have used the LDV device to measure the flow characteristics in different areas, such as, polymer solution, laminar flow and turbulent flow; and to study the LDV performance from the theoretical stand point.

In general, a LDV system can be considered as made up of two parts, the optical system and the signal processor. The optical system consists of the following components, Figure 2: (1) a laser which is used to provide a main focused light source, (2) a beam splitter which is used to separate the initial beam into two equally intensified beams, (3) focusing lens are forcing two beams to cross at point, and forming fringe pattern at the crossing point, (4) collecting lens are used to collect light into a photodetector, and (5) photodetector is used to convert the light energy into the electrical energy. The fringe pattern, or measuring volume, formed by intersection of two beams is the key zone which makes LDV function. When a particle moving with a fluid impinges the measuring volume, the light is scattered in all directions and some of it is picked

up by a photodetector. The movement of particle shifts the frequency of the scattered light by an amount known as the "Dopple shift" which is proportional to a component of the particle velocity. By knowing the fringe space the particle velocity can then be calculated. The fringe space is the function of intersection angle, θ , and wave length, λ , of the incident beams, and can be expressed as the following equation,

$$a = \frac{\lambda}{2 \sin \frac{\theta}{2}} .$$

In order to obtain good signal from a LDV system, the optical system must first of all be properly aligned.

Signal processor is a part of LDV device which converts the Doppler frequency into a readable voltage signal, from which the frequency can be calculated. An oscilloscope can also be used as an auxilliary device connecting to the signal processor to observe frequency signal, and the frequency signal can be compared with the digital value obtained from the signal processor.

In operating the signal processor the approximate particle velocity has to be known before-hand, so that the high pass and low pass filters can be set at the neighborhoods of the particle velocity and filt out "noise" frequency. The noise frequency is possibly contributed by room lights, photodetector, and the variation of particle velocity as it goes through the measuring volume.

Signal processor can be categorized into two types: Counter

and Tracker. Counter has the high/low pass filter accessory and it is required to know the measurement velocity in advance in order to set the high/low pass filter at the nearby measurement frequency. For using Tracker one has to, first of all, find or "track" a signal, and as the signal is "locked" properly the Tracker will adjust automatically in 15% of frequency changes. In this study since the velocity of the center line and the cross section is changing, it is difficult to pre-estimate the velocity at each new location. Therefore, Tracker is used as the primary signal processor to obtain the frequency, and Counter is used as the secondary unit to compare the results.

II. Experimental Apparatus

(a) Flow System

A 22-inch long and 4-inch ID plastic circular tube is used as a testing section for this study, Figure 1, the thickness of the tube is about $\frac{1}{4}$ ". Annulus plates with 1-inch thick and 2-inch inside circular hole are co-axially attached to the 22-inch long testing tube which generates a sudden expansion flow pattern inside the tube. The apparatus is to simulate the ramjet combustion system. The back end of the tube is connected to vacuum cleaners and the flow is sucked in from the front end of the tube. A capillary tube is located at the inlet of the testing section, and the capillary is connected to a manometer to read the inlet air velocity. Two pieces of co-axial tubes are connected to the both ends of the testing section tube.

The front end is used to regulate the aerosol particles flow and the back end is used to eliminate the end effect of exit flow.

The entire tube is fixed, and the laser optical system can be moved axially and radially so that the intersection of two beams - fringe pattern or measuring volume - can be adjusted at any axial and radial locations. Baking soda is used as solid particles and the particle diameter is about $0.5\text{-}\mu$ in diameter.

(b) Optical System

The major characteristic of optical system is discussed in the previous section and is also depicted in Figure 2. An item which has not been discussed is Frequency Shifter. The frequency shifter is an acousto-optic cell (Bragg) which is used to shift the laser light frequency. By properly shifting the laser light frequency one will be able to set the high/low pass filter of Counter Signal Processor at constant levels during a sequence of measurements, and also can measure the reversible flow velocity. For example if one is going to measure the Doppler shift frequency changing somewhere from 10MHz to 2 MHz, and by using frequency shifter of 5MHz with the flow velocity, then the Doppler shift frequency will be about 15 MHz to 7 MHz. Therefore, one may decide to set the high/low pass filter at 32 MHz and 8 MHz during the sequence of measurements.

III. Results

The profiles of velocity and turbulent intensity along the center line of the testing tube are depicted in nondimensional quantities on Figures 3 and 4. The figures show that the velocity decreases to about 25% of the inlet velocity when X/D approaches to about 10 where the flow is fully developed. The turbulent intensity increases from zero at the inlet point to the maximum at X/D equals to about 6-8, and decreases down to the steady state region. At the high turbulent intensity region the scatter of data is large, especially for the intensity data, because at this region a lot of particles move fast and exceed the frequency which eventually the signal processor would lose to "track".

The velocity profile and turbulent intensity for the cross sectional areas are shown on Figures 5 to 8. At cross section of $X/D = 1.25$, it shows a negative velocity profile exists near the wall which indicates the separation of the boundary layer. Moon and Rudinger (10) show in their results that the boundary separation exists at about $X/D = 2.0$ which agrees with this study. The turbulent intensity for the cross sectional profile is extremely difficult to measure near the wall region, the complicated situation is probably contributed by the rebounding of particles from the wall which cause transversal motion of particles.

IV. Conclusion & Recommendation

The turbulent velocity profile of the center line and the

cross section for a sudden expansion of circular tube can be measured by using LDV technology with the regular signal processor - such as, Counter or Tracker. However, the turbulent intensity is still difficult to measure at the regions near the wall and where the turbulent intensity is high, it is because the velocity perturbation changes so randomly and rapidly at these regions. In order to overcome these problems, it is suggested to use a computer to collect and analyzed a large amount of data simultaneously.

Acknowledge

The author of this paper appreciates the technical assistances from the Ramjet Technology Group at WPAFB for completion of this work.

Nomenclature

a : fringe space

D : inlet diameter of testing tube, 2 inches

r : radial distance of testing tube

R : radius of testing tube, 2 inches

u : local velocity

U_c : center line velocity

U_{IN} : air inlet velocity

X : axial distance from inlet point

θ : intersection angle of laser beams

λ : wave length of laser beam

u' : local perturbation velocity

References

1. Yeh, Y., and Cummins, H. Z., "Localized Fluid Flow Measurements With an He-Ne Laser Spectrometer," Applied Physics Letters, 4, 176 (1964).
2. Goldstein, R. J., and Kreid, D. K., "Measurement of Laminar Flow Development in a Square Duct Using a Laser Doppler Flowmeter," J. Applied Mechanics, 34, 813 (1967).
3. Berman, N. S., "Flow Behavior of a Dilute Polymer Solution in Circular Tubes at Low Reynolds Numbers," AIChE J., 15, 137, (1969).
4. Denison, E. B., Stevenson, W. H., and Fox, R. W., "Pulsating Laminar Flow Measurements with a Directionally Sensitive Laser Velocimeter," ibid, 17, 781 (1971).
5. Angus, J. C., and Edwards, R. V., "Signal Broadening in the Laser Doppler Velocimeter," ibid, 17, 1509 (1971).
6. Goldstein, R. J., and Hagen, W. F., "Turbulent Flow Measurements Utilizing the Doppler Shift of Scattered Laser Radiation," Physics of Fluids, 10, 1349 (1967).
7. Greated, C., "Measurement of Turbulence Statistics with a Laser Velocimeter," J. of Physics E, 3, 158, (1970).
8. Greated, C., "Measurement of Reynolds Stresses Using an Improved Laser Flowmeter," ibid, 3, 753, (1970).
9. Durst, F., Melling, A., and Whitelaw, J. H., "Low Reynolds Number Over a Plane Symmetric Sudden Expansion," J. Fluid Mechanics, 64, 111, (1974).
10. Moon, L. F., and Rudinger, G., "Velocity Distribution in an Abruptly Expanding Circular Duct," Trans. of the ASME, March, 226, (1977).

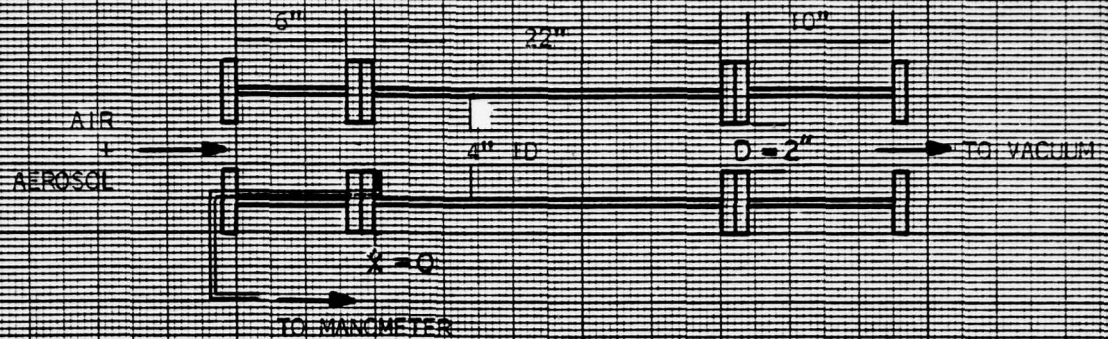


FIG. 1: FLOW SYSTEM

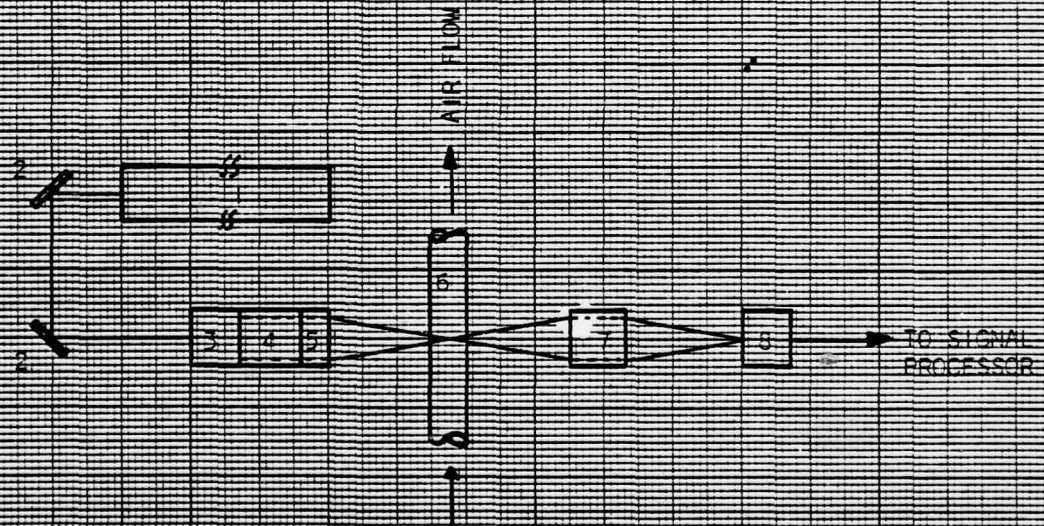


FIG. 2: OPTICAL SYSTEM

- | | |
|------------------------|--------------------|
| 1: ARGON IONIZED LASER | 5: FOCUSING LENS |
| 2: REFLECTION MIRROR | 6: CIRCULAR TUBE |
| 3: BEAM SPLITTER | 7: COLLECTING LENS |
| 4: FREQUENCY SHIFTER | 8: PHOTO DETECTOR |

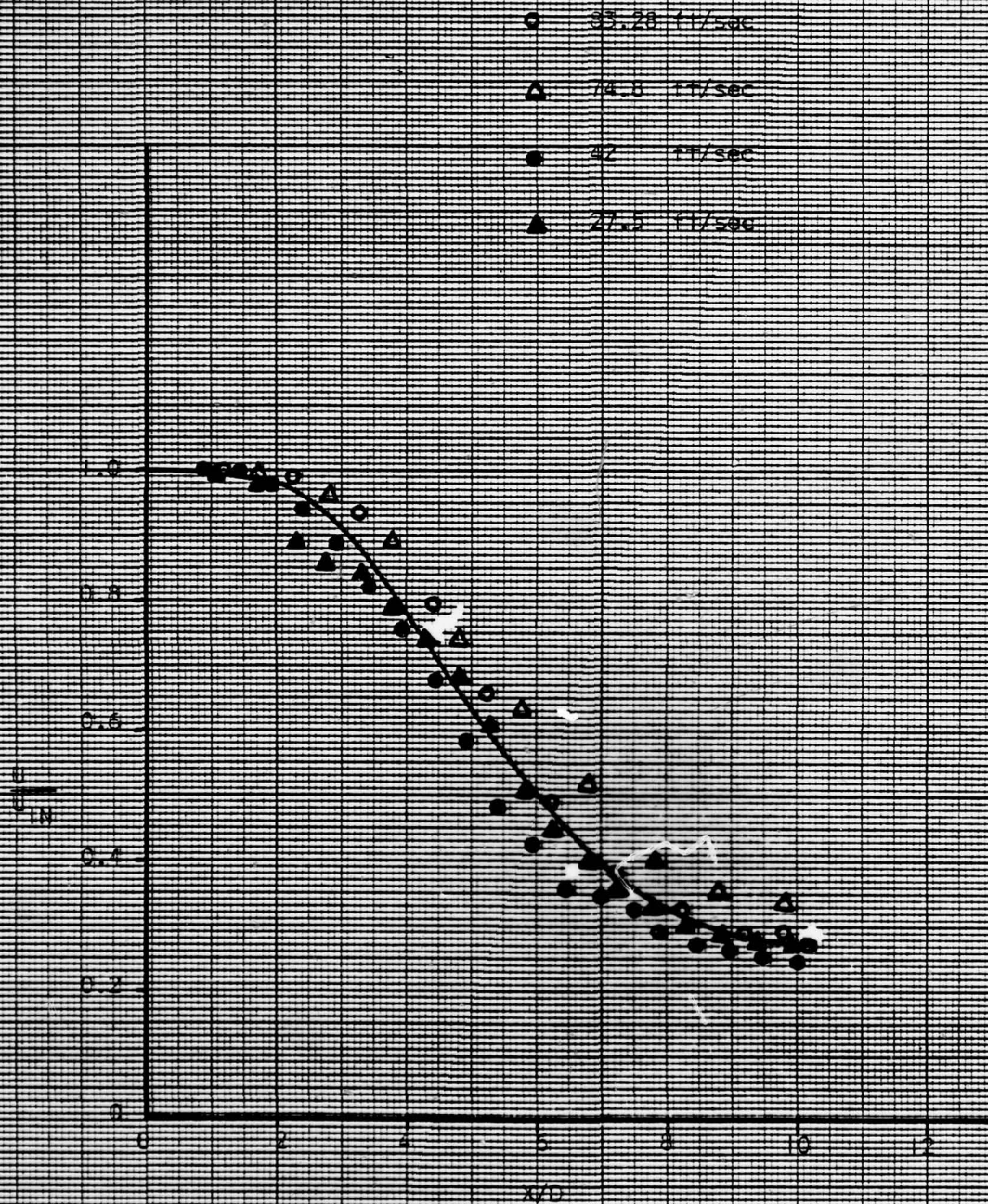


FIG. 5: VELOCITY PROFILE AT CENTER LINE

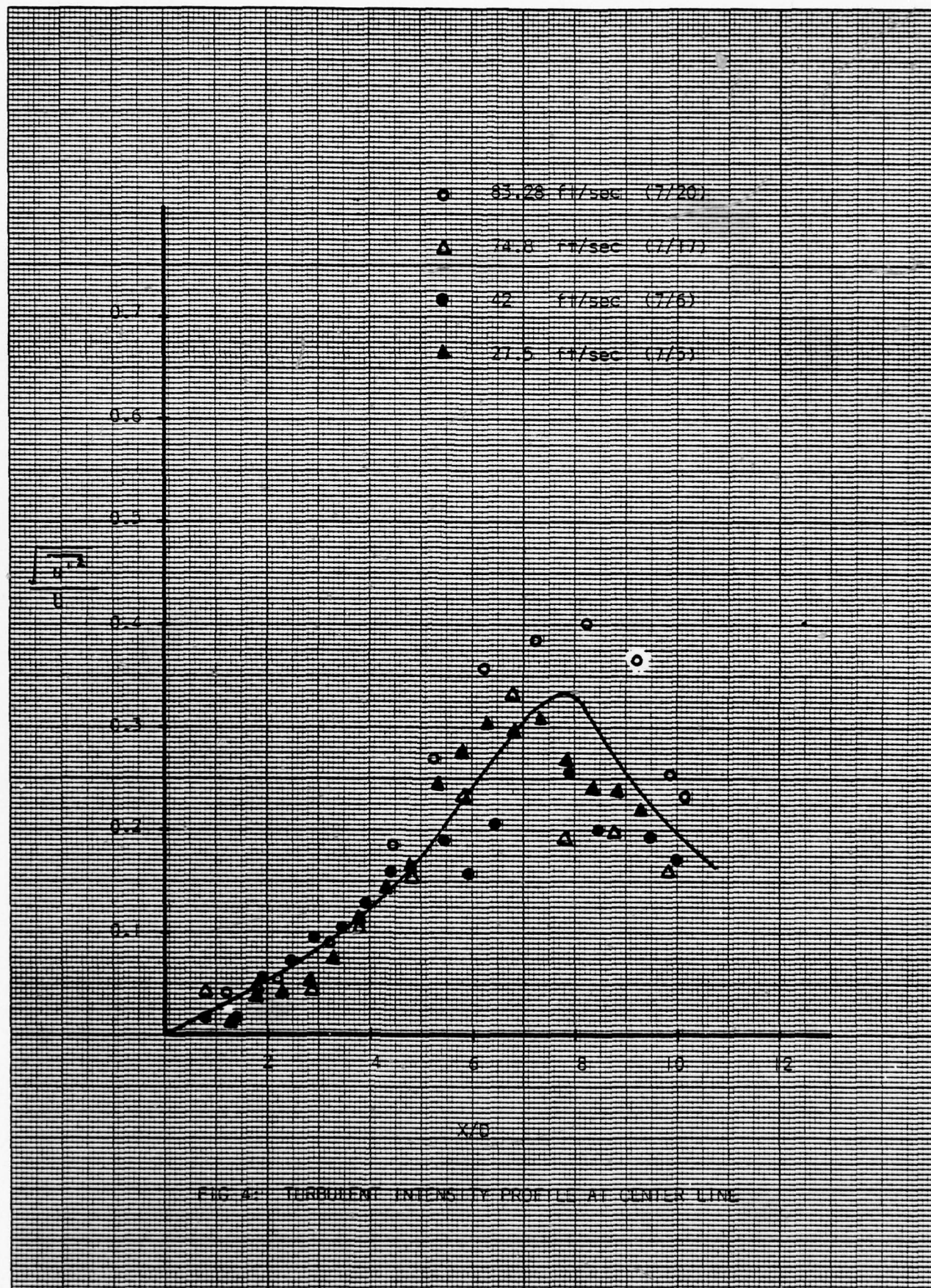


FIG. 4- TURBULENCE INTENSITY PROFILES AT CENTER LINE

1 INCH DIAMETER NCC
0.06 IN

NE IRM TUBE APH 2
MILLIMETER

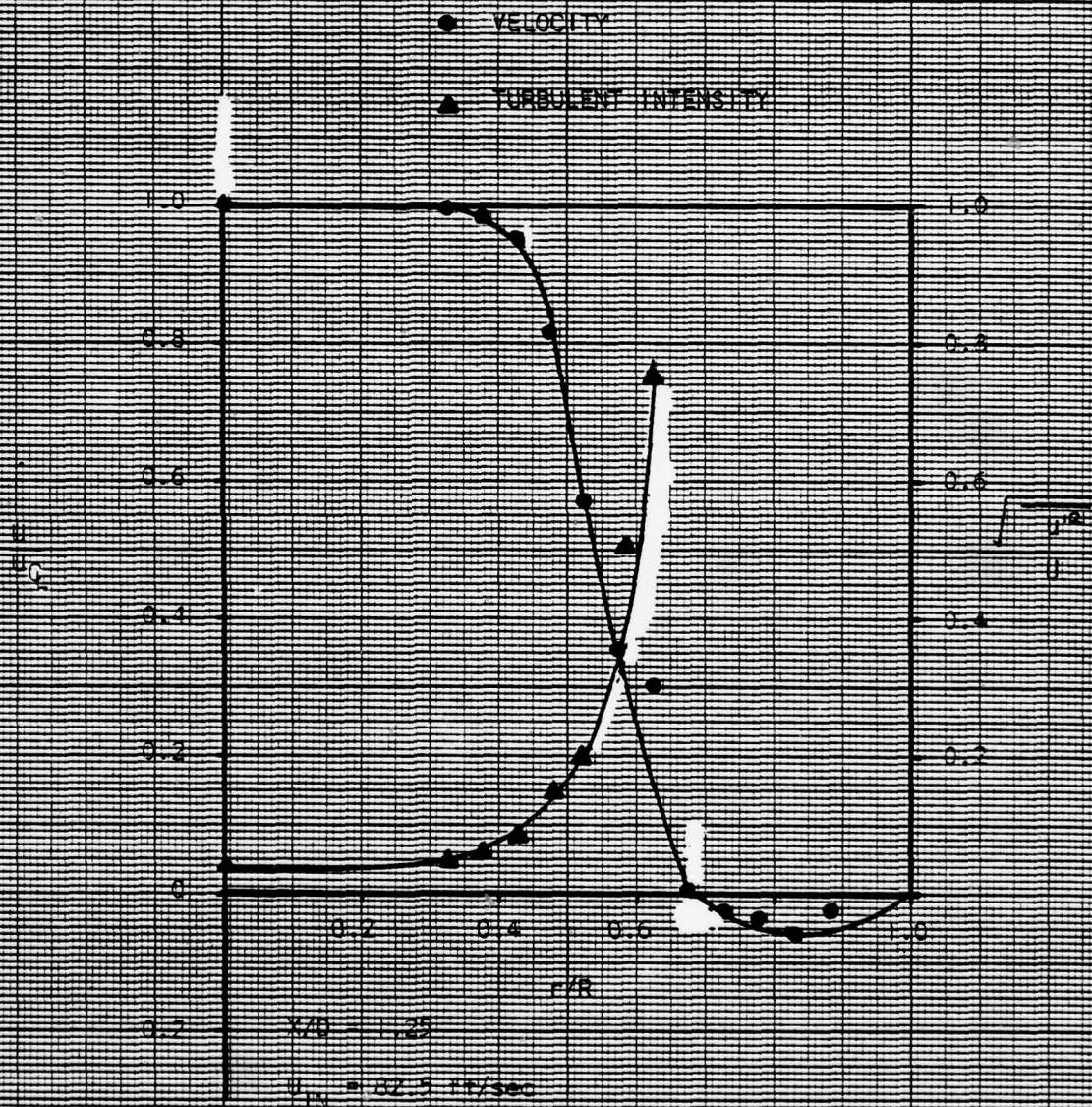
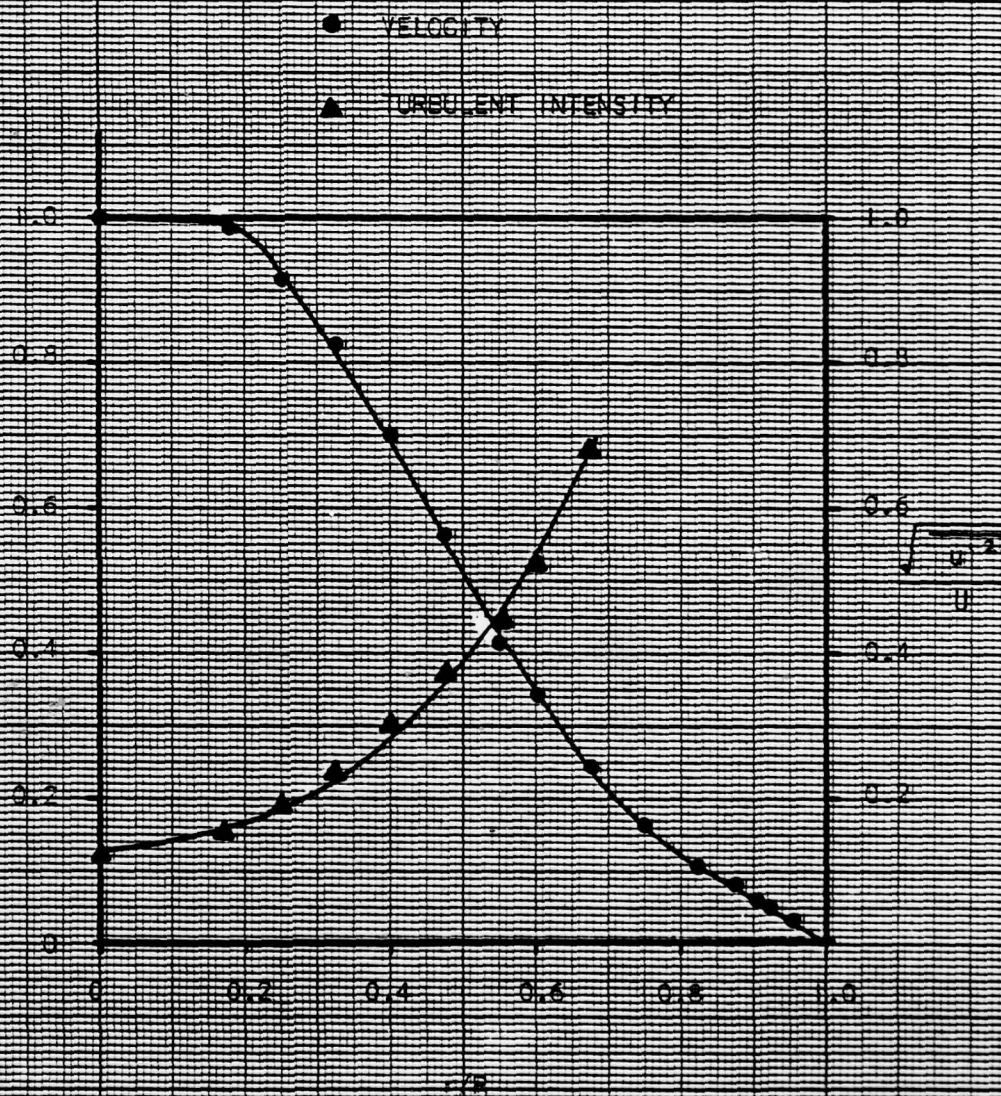


FIG. 5: VELOCITY & TURBULENT INTENSITY FOR CROSS SECTION



$$X/D = 5.75$$

$$U_{IN} = 32.45 \text{ ft/sec}$$

FIG. 6- VELOCITY & TURBULENT INTENSITY FOR CROSS SECTION



FIG. 7A: VELOCITY & TURBULENCE INTENSITY FOR CROSS SECTION

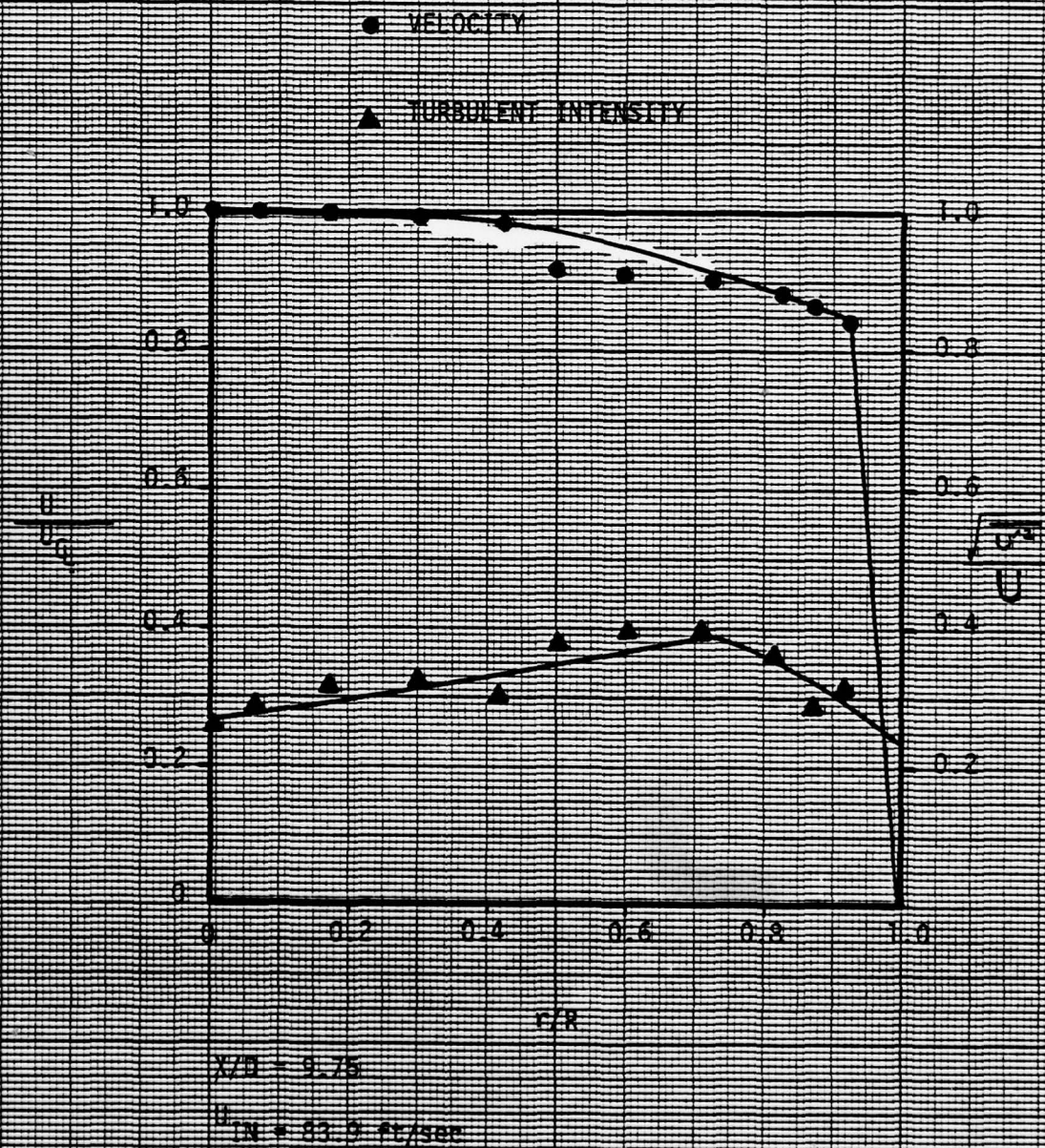


FIG. 8: VELOCITY & TURBULENT INTENSITY FOR CROSS SECTION

PRELIMINARY DESIGN PROCEDURE FOR HIGH POWER DENSITY MHD GENERATORS

PAU-CHANG LU*

Abstract

The steps to be taken in the preliminary design of a high power density MHD generator are formalized. The recommended design procedure starts with the optimum choice of the combustion chamber pressure. The optimization is based on a semi-empirical expression of the effective power developed recently by Smith and Nichols. As a result of this optimization, a rough estimate of the realizable power density is made, which yields the order of magnitude of the transverse dimension (keeping the length-to-diameter ratio around 10) for the desired power output. It is then recommended that the area variation (and length) be calculated on the basis of an isothermal core flow. This preliminary shape of the duct will serve as the base on which variations can be made by a computer to accommodate wall effects, in the detailed design stage. An alternative route for the preliminary design is also provided by using scaling laws. Starting with a well-designed generator which is demonstrably high in power density, dynamically similar units can be produced using these laws. The laws are developed following the modern procedure of ordering. Numerical examples are provided to illustrate the procedure.

* Professor of Mechanical Engineering, University of Nebraska-Lincoln, USAF-ASEE Summer Faculty Research Fellow, Wright-Patterson Air Force Base, Ohio.

FOREWORD

This is the final report on Preliminary Design of High Power Density MHD Generators under the USAF-ASEE Summer Faculty Program (WPAFB), administered at the Ohio State University, from June 5 to 16, and from June 26 to August 18 1978.

Dr. J. F. Holt of the Aero Propulsion Laboratory suggested the present study. The author appreciates greatly the opportunity of engaging Dr. Holt in enlightening discussions on numerous occasions.

TABLE OF CONTENTS

SECTION		PAGE
I	INTRODUCTION	1
II	SELECTION OF PRESSURE LEVEL	5
III	ISOTHERMAL DESIGN	12
IV	SCALING LAWS	17
V	LOSSES	29
	References	31
	List of Symbols	33

LIST OF ILLUSTRATIONS

FIGURE		PAGE
II-1	Sketches of w_{eff} versus β	8
III-1	A Duct with Diagonal Conducting Walls	13

LIST OF TABLES

TABLE		PAGE
II-1	Optimal Combustion-Chamber Pressure for Stoichiometric Combustion of Toluene and Oxygen Seeded with Cesium Carbonate (4 T)	10
IV-1	Design via Scaling Laws	27

SECTION I

INTRODUCTION

Designing MHD generators for high power density is currently limited to a small circle of practitioners, away from the main stream of MHD activities; the planning is usually done in an ad hoc manner, for individual cases, without stating clearly the approach followed or the philosophy adopted. To render such planning more a science than an art, it is the purpose of the present study to formalize the steps to be taken in the preliminary design of a high power density MHD generator. These recommended steps are gathered here (presumably for the first time) for one object (and one object only): the realization of maximum possible power generated per unit volume. Detailed design calculations (on a computer) that follow these preliminary steps will undoubtedly indicate trade-off points for a variety of meritous features. But in this report, only maximum power density is being pursued.

In gathering material from a widely scattered literature for the expressed purpose, the author can hardly claim any originality. Although critical comments, personal judgments, minor discoveries, slight extensions and small variations abound, this report remains but a designer's guidebook.

In Section II, the recommended design procedure starts with the optimum choice of the combustion chamber pressure, after a brief description of optimization calculations to be done on the inlet Mach number, seeding ratio, and O/F (oxygen to fuel ratio). The optimization is based

on a semi-empirical expression of the effective power output developed only recently by Smith and Nichols (ref. 1). This new approach apparently provides the most rational basis to the design procedure to date.

As a result of Section II, a rough estimate of the realizable power density emerges, which will yield the order of magnitude of the transverse dimension (keeping the length/diameter around 10) for the desired power output.

In Section III, it is recommended that the area-variation (and the length) of the generator be estimated on the basis of an isothermal core flow. The rationale here is as follows: The plasma conductivity varies exponentially with temperature; therefore, keeping the entire duct uniformly at a high temperature level would promote high power density. (In contrast, a constant velocity design would incur heavy temperature drop and pressure loss; the latter would also burden the diffuser heavily.) This preliminary shape of the duct will eventually serve as the base on which variations will be made on a computer to accomodate the wall effects. The final (computer-aided) design, of course, will not come out isothermal; but its temperature variation in the flow direction will certainly be relatively small.

Section III contains formulas developed for the isothermal core flow through a diagonal conducting-wall generator. To the author's present knowlegde, these extended formulas for the case of diagonal conducting walls are new.

In Section IV, the preliminary design is carried out along a different route. Starting with any well designed generator which is

demonstrably high in power density, either already in operation or in an advanced stage of planning, scaling laws can be applied to produce dynamically similar units for a different power output, and/or magnet strength, and/or fuel, etc. The procedure can also be used to yield a dynamically similar pilot unit which can be tested before embarking on the the larger-sized endeavor. It must be emphasized here that, if a unit has a power density which is maximum under the given restraints, its dynamically similar models will deliver far smaller power per unit volume (being still "high", possibly).

The key modeling parameter involved in the scaling laws are established in Section IV following a modern procedure known as the ordering process (see, e.g., Chapter 5 of ref. 2). Modeling (or scaling) is then carried out in the classical manner (see, e.g., Chapter 4 of ref. 3). Although the resulting laws are identical with those quoted in the literature (e.g., Chapter 7 of ref. 4), the derivation presented in this report is eminently more convincing.

Finally, in the short Section V, minor losses near the walls are discussed. The discussion is brief since these losses will be accounted for , anyway, in the next step of the design--the computer-aided simulation and selection.

It is hoped that generator designers will find the recommended procedure helpful in providing initial inputs to the design of magnets (which usually has to be started simultaneously with that of generators), as well as to the sophisticated numerical programs (the final design tools)

which are available today.

SECTION II

SELECTION OF PRESSURE LEVEL

A modern and operational definition of design refers to it as "optimization under partially uncertain constraints." With this definition in mind, one may state a general design philosophy or approach in the form of five steps: (1) Assuming that the operation is not too sensitive to changes of various parameters in a rather large neighborhood of the optimal condition, optimize the object quantity with respect to the parameters one after the other. (2) Establish a rational guideline for the optimization with respect to each parameter. (3) Display a number of optimal calculations over a range of uncertain values of the constraining parameters. Select a few (or one), exercising the designer's judgement. (4) Trade off (i.e., deviate from the optimal) for other desired or required characteristics. And, (5) simulate the few cases which exhibit overall possibilities on a computer, and make a final decision.

With reference to the present task of designing a MHD generator for high power density, we realize immediately that the object quantity to be maximized in the above steps is the power output per unit volume w . The parameters to be considered at the outset are the fuel used, the oxydizer, the seeding material, the seeding ratio, the O/F ratio, the inlet Mach number M_1 , and the combustion chamber pressure p^0 . (All products of combustion refer to hydrocarbon fuels.)

As a measure of the effective power density when loaded for

maximum power, we will adopt the following semi-empirical formulas due to Smith and Nichols (ref. 1):

Faraday (Segmented)--

$$w_{eff} \sim \begin{cases} \left(\frac{1 + \beta^*}{1 + \beta^{*2}} \right) \left(\frac{1}{4} \right) \sigma^* u^{*2} B^{*2}, & \beta^* > 1 \\ \left(\frac{1}{4} \right) \sigma^* u^{*2} B^{*2}, & \beta^* \leq 1 \end{cases}$$

Diagonal Conducting Walls--

$$w_{eff} \sim \begin{cases} \frac{(\alpha^* \beta^* - 1)^2}{\beta^* (1 + \alpha^{*2}) (1 + \beta^{*2})} \left(\frac{1}{4} \right) \sigma^* u^{*2} B^{*2}, & \beta^* > 1 \\ \frac{(\alpha^* \beta^* - 1)^2}{(1 + \alpha^{*2}) (1 + \beta^{*2})} \left(\frac{1}{4} \right) \sigma^* u^{*2} B^{*2}, & \beta^* \leq 1 \end{cases}$$

where the asterisk is used to indicate evaluation of quantities at certain reference point (the inlet, for example), and where

β = Hall parameter

σ = conductivity

u = flow velocity

B = magnetic field strength

α = (Hall field)/(Faraday field)

In an empirical (and approximate) manner, the formulas account for the internal current leakage and electrode voltage drops. (This empirical aspect probably also prevents the first formula from being a special case of the second.)

For a given fuel mixture, temperature level, and B^* , the w_{eff}

vs. β^* curve (remembering that $\sigma^* \propto \sqrt{\beta^*}$ because of the pressure variation) shows a trend as sketched in Fig. II-1 (the curve marked diagonal conducting walls being roughly for $\alpha = -\beta$, a power-maximizing value). It is thus observed that w_{eff} is maximum when $\beta^* = 1$, and that w_{eff} decreases much faster for decreasing β^* below 1 than increasing β^* above 1.

A rational guideline for the selection of the pressure level, represented by p^* , now emerges: Choose p^* such that β^* , for the given T^* and B^* , equals 1; and, if practical constraints force a deviation, make $\beta^* > 1$ (rather than < 1).

In following this guideline, there is still the question of where to enforce it. Should the asterisk refer to the nozzle exit, where the magnetic field peaks, or some kind of average state? Anticipating an isothermal (or, approximately isothermal) design, σ will stay the same from the inlet to the peak point of the field. Also, from existing designs where temperature drop is relatively slight, we observe that the flow velocity hardly changes before the peak of the magnetic field. Therefore, it seems desirable to select the pressure value so as to make $\beta = 1$ (or slightly above) at the peak of the field. Thus, with a computer print-out of $\beta(T, p, B)$ in hand (actually, $\beta \propto B$), one will select this pressure level accordingly.

Once the pressure at the peak point is chosen, one can add an empirical percentage ($\sim 10\%$) to it to obtain p_1 at the inlet. (In practice, this amounts to a minor option only, since a preliminary

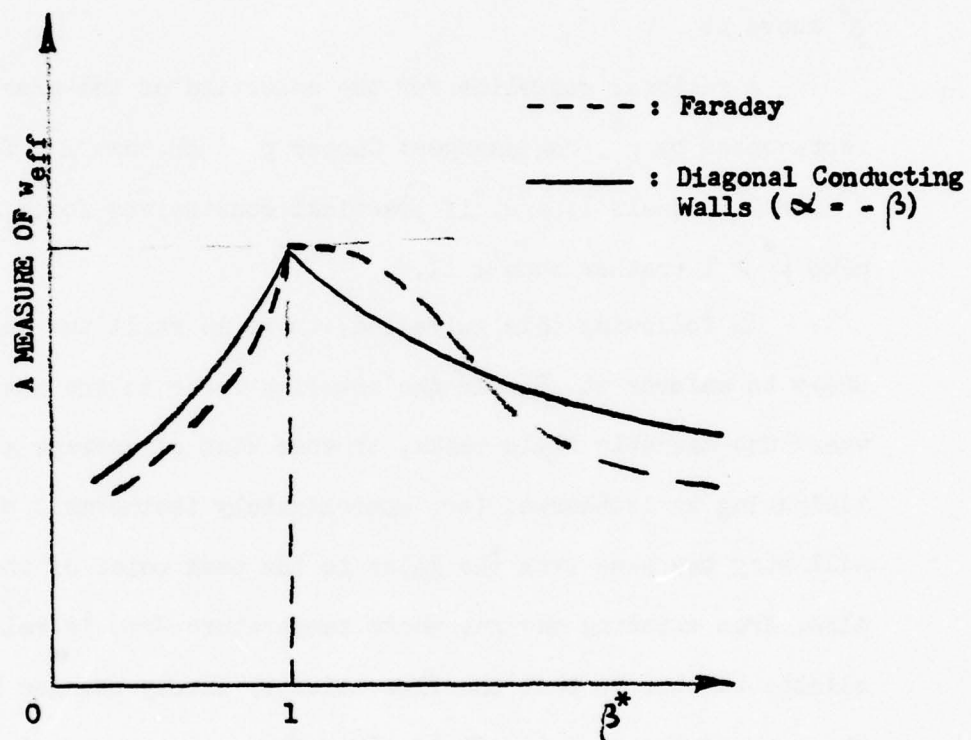


Figure II-1 Sketches of w_{eff} versus β .

design is routinely within a band of variation wider than 10%.) After this, the combustion pressure can be estimated for every given M_1 (knowing the equivalent ratio γ of the specific heat capacities of the plasma). Incidentally, M_1 also links the generator temperature with the combustion temperature T^0 .

As an example, toluene + O_2 + Cs_2CO_3 at the stoichiometric O/F ratio is treated in the manner just described for a peak magnetic field of 4 T. The result is summarized in Table II-1. Two effective combustion temperatures are employed in the table since the temperature level at the nozzle entrance may be raised somewhat by an increase in combustion pressure, and/or an improvement in the combustion chamber design, etc.

It must be emphasized once again that the optimal pressure level is chosen independently of optimization with respect to the other operational parameters. If all parameters are optimal, the effective power density will be the maximum of maxima; otherwise, it will only be the maximum for a given (non-optimal) set of parameter values. As a matter of fact, in Table II-1, although the Mach number is in the optimal range, the seeding and the O/F ratios are both rather far from being optimal.

With the aid of a computer, the effective power density can be optimized with respect to each one of the four parameters, seeding ratio, O/F ratio, inlet Mach number, and the pressure level. Essentially, the computer is to print out a chart showing the variation of w_{eff} with respect to these parameters; the optimal values are

TABLE II-1

OPTIMAL COMBUSTION-CHAMBER PRESSURE FOR STOI-
 CHIOMETRIC COMBUSTION OF TOLUENE AND OXYGEN
 SEEDED WITH CESIUM CARBONATE
 (4 TESLA)

10% Seeding By Weight Of Fuel			30% Seeding By Weight Of Fuel	
M_1	$T^0 = 3100 \text{ K}$	3400 K	3100 K	3400 K
1.9	$p^0 = 12 \text{ atm}$	14 atm	7.5 atm	10 atm
2.0	14 atm	16.5 atm	8.5 atm	11 atm
2.1	16 atm	18 atm	10 atm	13 atm
2.2	19 atm	22 atm	12.5 atm	14.5 atm

then easily identified. Actually, less elaborate study (ref. 5) has shown that M_1 should always be around 2 for products of combustion to realize maximum power density (which fact is also borne out by a detailed numerical example in ref. 1). Thus, for a preliminary design, anticipating certain practical ranges of the seeding and O/F ratios (which should be, but might not be, around the optimal value) and M_1 (which must be around 2), the selection of pressure level may as well be effected by following the suggested guideline as exemplified in Table II-1.

Finally, each selection in Table II-1 has an anticipated power density associated with it; we will quote four numbers here as illustrations:

10% seeding, $M_1 = 2.1$ --

$T^0 = 3100 \text{ K: } w \sim 60 \text{ MW/m}^3$

$T^0 = 3400 \text{ K: } w \sim 320 \text{ MW/m}^3$

30% seeding, $M_1 = 2.1$ --

$T^0 = 3100 \text{ K: } w \sim 125 \text{ MW/m}^3$

$T^0 = 3400 \text{ K: } w \sim 620 \text{ MW/m}^3$

From these figures, the generator volume can be estimated for a desired power output. In addition, if an empirical length-to-diameter ratio (~ 10) is adopted, based on a compromise between end and wall effects, the linear size of the generator can also be estimated.

SECTION III

ISOTHERMAL DESIGN

After the optimization discussed in the previous section, the design philosophy calls for the determination of the duct shape (i.e., the area variation in the flow direction), as far as the core is concerned, for isothermal generation of electricity. (The result will serve as the base shape upon which the wall effects will be added in the computer simulation that follows the preliminary design.) To this end, let us first collect all the governing equations (referring to Fig. III-1) which describe a general core flow:

$$\rho u A = \text{constant} (= \rho_1 u_1 A_1) \quad (1)$$

$$\rho u u' = -p' + j_y B \quad (2)$$

$$\rho u \{c_p T' + (u^2/2)'\} = j_x E_x + j_y E_y \quad (3)$$

$$\rho = p/RT \quad (4)$$

$$E_x = \alpha E_y \quad (5)$$

$$K = E_y / uB \quad (6)$$

$$j_x = \{\sigma / (1 + \beta^2)\} (uB) \{\alpha K - \beta(K - 1)\} \quad (7)$$

$$j_y = \{\sigma / (1 + \beta^2)\} (uB) \{\alpha \beta K + (K - 1)\} \quad (8)$$

where prime denotes differentiation with respect to x ; subscripts x and y indicate specific components; ρ , A , j , E , R , c_p , and K are respectively the plasma density, cross-sectional area of flow passage, current density, electric field, gas constant of the plasma, specific heat capacity at constant pressure of plasma, and the loading factor; and, all quantities are local. In Equation (1), ρ_1 and u_1 are known from the

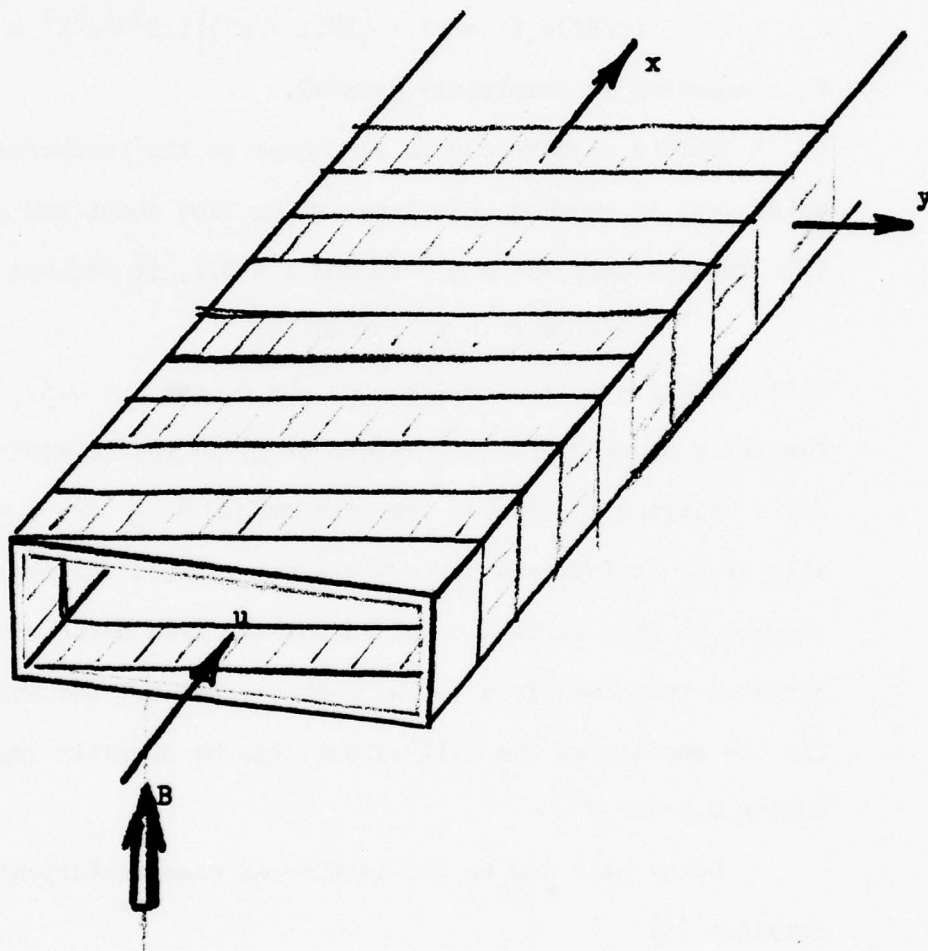


Figure III-1 A Duct with Diagonal Conducting Walls.

procedure presented in the previous section; A_1 can also be decided roughly on the power desired as explained at the end of the previous section.

In the above, Equations (2) through (8) can be easily combined into the following key equation:

$$(p/RT)c_p T' = p' + \left\{ \sigma / (1 + \beta^2) \right\} (uB^2) \left\{ \alpha^2 K^2 + (K - 1)^2 \right\} \quad (9)$$

This equation is completely general.

Before we narrow down our scope to the isothermal case, let us point out, in passing, one interesting fact about the general Equation (9): For the case where $\alpha = -\beta$ and $K = 0.5$, it reduces to the form

$$(p/RT)c_p T' = p' + O(uB^2/4) \quad (10)$$

which, being exactly the case for $\beta = 0$ (and $K = 0.5$), has been extensively studied (for all values of K) in the literature (see ref. 6 for a unifying approach). The case with $\alpha = -\beta$ and $K = 0.5$ is practically of great interest since $\alpha = -\beta$ maximizes the power density (with respect to α), while $K = 0.5$ is not far from being optimal unless deviates very much from 1. For such a case, all the available solutions (in the absence of the Hall effect) can be directly employed by substituting 0.5 for K .

Going back now to the isothermal case exclusively, we have from Equation (9)

$$p' = - \left\{ \sigma / (1 + \beta^2) \right\} (uB^2) \left\{ \alpha^2 K^2 + (K - 1)^2 \right\} \quad (11)$$

But, from Equation (2) and (8),

$$\left\{ \sigma / (1 + \beta^2) \right\} (uB^2) = \left\{ \rho u u' + p' \right\} / \left\{ \alpha \beta K + (K - 1) \right\} \quad (12)$$

Thus, combining Equations (11) and (12), we have

$$p' = -\{\alpha^2 K^2 + (K - 1)^2\} \{p u u' + p'\} / \{\alpha \beta K + (K - 1)\} \quad (13)$$

or,

$$K\{K(\alpha^2 + 1) + (\alpha\beta - 1)\}p' = -\{\alpha^2 K^2 + (K - 1)^2\}(p/RT)(u u')$$

Integrating, we obtain finally the relation (for constant α , β , and K):

$$p/p_1 = \exp \{a(u_1^2 - u^2)\} \quad (14)$$

where

$$a = \frac{\alpha^2 K^2 + (K - 1)^2}{(2RT)K\{K(\alpha^2 + 1) + (\alpha\beta - 1)\}} \quad (15)$$

For a variable $\beta (= b/p)$, but with $\alpha = -\beta$ and constant K , we have

$$(u_1^2 - u^2)/2RT = \ln \left\{ \left(\frac{p}{p_1} \right)^{\frac{K-1}{K}} \left(\frac{p_1^2(K-1)^2 + bK}{p^2(K-1)^2 + bK} \right)^{\frac{2K-1}{2K(1-K)}} \right\} \quad (16)$$

After establishing such a p vs. u relationship, we can go back to Equation (2) and integrate it with respect to x . The result, with the help of Equation (13), is as follows:

$$x = - \int_p^{p_1} \frac{(1 + \beta^2) dp}{\{\alpha^2 K^2 + (K - 1)^2\} (\sigma u B^2)} \quad (17)$$

where the p vs. u relationship is to be incurred in the integration.

Note also that Equation (17) is completely general, with varying σ , β , etc.

In any given case, $p(x)$ obviously will become known at this step; $u(x)$ will then emerge from the p vs. u relationship. Next,

since $\rho \propto p$, $A(x)$ can be calculated via Equation (1). To determine the total length of the generator, one simply sums up the power output along the duct until the desired total output is reached.

(Although it is not likely that the exit pressure would turn out lower than 0.2 atm; but, nonetheless, this value must be checked. An exit pressure lower than 0.2 atm will burden the diffuser heavily.)

Based on such a preliminary core design, end effects, wall losses, as well as real-gas effects can be added to yield the adjusted duct shape. It is not known whether the above "almost-isothermal" procedure is actually being used by the designers in the field. But it is gratifying to note that in at least two real cases (both supersonic and for high power density) the temperature deviation is kept below 2.5%: it decreases by about 2.3% in the AVCO 400 kW generator (ref. 7), and increases by 1.3% in the Russian PAMIR I generator (ref. 8).

SECTION IV

SCALING LAWS

The word "scaling" as used here is restricted to mean dynamic and geometric similitudes among a series of MHD generators. The aim of scaling is to ensure that, once one member of the series is probed extensively (either by actual measurements if it is already operating, or by detailed numerical simulation if it is not yet built), the performance of any other member of the series can be predicted.

The fundamental tool behind scaling is dimensional analysis (or, alternatively, ordering). In principle, every physical phenomenon can be described by a list of dimensionless parameters (in addition to geometric shape, ratios, and angles). Two specific instances of that phenomenon are mutually transformable in a dynamically similar manner if they have the same numerical values for all the modeling parameters, with geometric similarity taken for granted. Scaling, or modeling (see Chapter 4 of ref. 3), then calls for keeping the dimensionless parameters the same among individual members of a series.

In practice, if the dimensionless parameters governing a certain phenomenon are many, a dynamically similar series may very well end up containing only one member. The conclusion of the scaling process then becomes degenerate and trivial--namely, every specific instance is only similar to itself. This would be the end of scaling, requiring individual instances be probed (numerically or experimentally) separately.

In order to have a meaningful correlation among different mem-

bers of a series, certain dimensionless parameters may have to be allowed to vary from member to member. If the parameters left open in this manner are the geometric ratios and angles, we have a distorted scaling; otherwise, we have a partial scaling.

For MHD generators, a (severely) distorted scaling will require detailed numerical simulation to compensate for the drastic deviations in the predicted performance. Since scaling is basically a tool used to provide quick guidelines before detailed analysis, a distorted scaling is of no use to MHD generator design--one might as well skip to numerical simulation directly.

The rest of this section, therefore, is devoted exclusively to the partial scaling of MHD generators.

Applying the ordering process (see Chapter 5 of ref. 2) to Equations (1) through (8) (and thereby ignoring end and wall effects), we introduce the following representative quantities:

Density---- ρ_1

Velocity---- u_1

Area---- A_1

x-coordinate---- Generator length L (which is proportional to $\sqrt{A_1}$ because of geometric similitude)

Magnetic field---- Peak magnetic field B^*

Pressure---- p_1

Temperature---- T_1

Conductivity---- σ_1

Current density----- $\sigma_1 u_1 B^*$

Electrical field----- $u_1 B^*$

Substituting into these equations the dimensionless quantities $\tilde{\rho} = \rho/\rho_1$, etc., with the tilde signifying nondimensionalization, we have

$$\left\{ \begin{array}{l} \tilde{\rho} \tilde{u} \tilde{A} = 1 \\ \gamma M_1^2 (\tilde{\rho} \tilde{u} \tilde{u}') = -\tilde{p}' + S \tilde{j}_y \tilde{B} \\ \tilde{\rho} \tilde{u} \left\{ \tilde{T}' + [(\gamma - 1)/2] M_1^2 (\tilde{u}^2/2)' \right\} = [(\gamma - 1)/\gamma] S (\tilde{j}_x \tilde{E}_x + \tilde{j}_y \tilde{E}_y) \\ \tilde{\rho} = \tilde{p}/\tilde{T} \\ \tilde{E}_x = \alpha \tilde{E}_y \\ K = \tilde{E}_y / \tilde{u} \tilde{B} \\ \tilde{j}_x = \left\{ \tilde{\sigma} / (1 + \beta^2) \right\} (\tilde{u} \tilde{B}) \left\{ \alpha K - \beta (K - 1) \right\} \\ \tilde{j}_y = \left\{ \tilde{\sigma} / (1 + \beta^2) \right\} (\tilde{u} \tilde{B}) \left\{ \alpha \beta K + (K - 1) \right\} \end{array} \right.$$

where the prime now denotes differentiation with respect to (x/L) . As a result of this process, a number of governing (dimensionless) parameters show up naturally and unambiguously; namely,

$$S = \sigma_1 u_1 B^{*2} L / \rho_1$$

$$M_1 = u_1 / \sqrt{\gamma R T_1}$$

$$\tilde{B} = \tilde{B}(x/L) = \tilde{B}(\tilde{x})$$

$$\tilde{\sigma} = \tilde{\sigma}(\tilde{T}, \tilde{p})$$

$$\beta = \left\{ \tilde{\beta}(\tilde{T}, \tilde{p}) \right\} \beta_1$$

$$\alpha, K, \gamma$$

A series of geometrically similar generators must have the same numerical values or variations (for \tilde{B} , $\tilde{\beta}$, and $\tilde{\sigma}$) in order to be dynamically similar, as far as the core is concerned. Out of this list, γ stays around 1.1 for all products of combustion; we can therefore omit

it from further consideration.

From the literature, we find that

$$\beta \propto B \sqrt{T}/p$$

and

$$\sigma \propto T^m/p^n$$

where m and n are universal constants (depending on the temperature range) for all plasmas (but where the proportionality constants differ for different plasmas). That is to say,

$$\tilde{\beta} = \tilde{B}(B^*/B_1) \sqrt{T_1}/p_1$$

and

$$\tilde{\sigma} = \tilde{T}^m/\tilde{p}^n$$

So, $\tilde{\sigma}(\tilde{T}, \tilde{p})$ retains the same form, only the consideration of σ_1 is needed. Similarly, although $\tilde{\beta}(\tilde{T}, \tilde{p})$ retains the same form (assuming fixed $\tilde{B}(\tilde{x})$), β_1 must be considered.

In practice, α is made close to $-\beta$; so, there is no necessity to include its variation. The loading factor K is either kept constant for a series of generators, or it is to vary only slightly; and, in addition, its influence on the power density can be estimated using a slug-flow model. Thus, finally the list of governing parameters is reduced to S , M_1 , $\tilde{B}(\tilde{x})$, and β_1 (with σ_1 embedded in S).

These governing parameters are also obtained, for example, by Garrison, Brogan, Nolan, et. al. (ref. 9), presumably through standard dimensional analysis. The parameter S also frequently appears in the literature by way of dimensional analysis; but, usually, it is the only

one mentioned.

Out of these four remaining quantities, $\tilde{B}(\tilde{x})$ is not likely to be fixed for a series of generators; we will have to start our partial scaling by ignoring the requirement of a fixed $\tilde{B}(\tilde{x})$. The influence of the field strength is then invested only with the peak B^* which appears in S (and in β_1 through B_1). Such a partial scaling has been carried out and applied by Garrison, Brogan, Nolan, et. al. (ref. 9). Although it is somewhat fruitful and useful, such scaling yields rather restrictive modeling laws.

Noting that the β -level in a well-designed generator is always around 1, and that the influence of the β -level can be assessed separately (e.g., in the manner discussed in Section II), this report will go one step further and recommend a partial scaling based on S and M_1 only. By relaxing the requirement that β_1 be kept constant from case to case in a series of generators, more productive scaling laws are possible; and the scope of application is suddenly widened. For instance, taking

$$\sigma = cT^{10}/p^{0.5} \quad (18)$$

where c is a coefficient the value of which is available for specific plasmas, we see that S = constant for different cases in a series implies that[#]

$$L \propto p^{1.5} B^{-2} T^{-10.5}/c \quad (19)$$

where M = constant, i.e.,

[#] From this point on, the subscript 1 and superscript * will be omitted for ease of writing.

$$u \propto T^{0.5} \quad (20)$$

(noting that the molecular masses for all products of combustion stay roughly around 35) is also enforced. Now, the object in building a MHD generator is to realize an electrical power output W which can be calculated by integrating the solutions of Equations (1) through (8) with respect to x . By examining the dimensionless forms of these equations, we conclude that W in the form of a dimensionless parameter must come out a function of (under the present partial scaling) S and M , without the necessity of actually solving the equations. To be more specific, we conclude that

$$W/(\sigma u^2 B^2 L^3) = F(S, M)$$

which yields, since S and M are both kept fixed,

$$W/(\sigma u^2 B^2 L^3) = \text{constant}$$

or,

$$\begin{aligned} W &\propto \sigma u^2 B^2 L^3 \\ &\propto c T^{11} B^2 L^3 / p^{0.5} \end{aligned} \quad (21)$$

or, in terms of the power density,

$$w \propto c T^{11} B^2 / p^{0.5} \quad (22)$$

Next, combining Equations (19) through (22) in various ways, we have the following additional scaling laws:

$$L \propto W^{0.375} / (c^{0.25} T^{2.75} B^{0.5}) \quad (23)$$

$$p \propto c^{0.5} T^{6.6} W^{0.25} B \quad (24)$$

$$w \propto B^{1.5} c^{0.75} T^{7.7} \quad (25)$$

where Equation (25) is approximate in the sense that $w^{1.125}/L^3$, instead of W/L^3 , is regarded as being proportional to w . Furthermore, Equation

(23) squared, multiplied by Equation (24), yields

$$W \propto pL^2/T^{1.1} \text{ (or, roughly, } pL^2 \text{)} \quad (26)$$

In applying these scaling laws, one must always bear in mind four things: (1) $\tilde{A}(\tilde{x})$ is fixed; (2) Equations (18) through (20) must be satisfied; (3) there are possible errors in ignoring $\tilde{B}(\tilde{x})$, α , β , and K ; and (4) deviations must be anticipated from end and wall losses. To guard against possible misapplications, let us quote here one counter-example: In applying Equation (26) to the same generator, one seems to see that W is proportional to p ; but this is singularly uninteresting, since Equation (19) then dictates that p be fixed (for fixed B , T , and c). A correct interpretation of Equation (26) would be, for example, thus: For a given generator, equipped with a different magnet, the pressure level must be adjusted according to Equation (19) in order to operate in a dynamically similar manner as when the old magnet is used; then (and only then), W will change in direct proportion to p . (If nothing but p is changed, the new operation will not be dynamically similar; and W will not be proportional to p .) We would also like to take this opportunity to emphasize the fact that it is the dimensionless quotient $w/\sigma u^2 B^2$ that remains fixed from member to member in a series of dynamically similar generators; the power density w definitely will change from case to case. It is especially important to bear in mind that, if a generator is designed for maximum power density, and if a scaled-down unit is first built, the smaller unit is definitely not going to be able to claim maximum

power density.

Among the above-quoted scaling laws, Equations (19), (23), (24), and (26) have been derived before by Rosa (Chapter 7 of ref. 4), using an intuitive, heuristic, and simplistic argument which is rather unconvincing (the requirement of fixed M being absent).

In the rest of this section, we will apply some of the scaling laws to scale up or down some existing designs known for their high power densities. We will use subscripts 1 and 2, respectively, to denote the base base design and the scaled unit. Temperature and pressure of the combustion chamber are used in the calculations; for fixed Mach number, they are proportional to those at the generator inlet. The two specific formulas used are Equations (19) and (22) which are rewritten as

$$L_2/L_1 = (c_1/c_2)(p_2/p_1)^{1.5}(B_1/B_2)^2(T_1/T_2)^{10.5} \quad (27)$$

$$w_2/w_1 = (c_2/c_1)(T_2/T_1)^{11}(p_1/p_2)^{0.5} \quad (28)$$

The results of the scaling are summarized in Table IV-1. (A part of the calculation was carried out by Dr. J. F. Holt of the Air Force Aero Propulsion Laboratory, the whom the author is indebted in many ways.) A detailed description of the base designs is given in the following:

Base Design #1--AFAPL KIVA-1 (ref. 10)

Toluene + O₂, seeded with Cs

p^o = 10 atm, T^o = 3100 K

M₁ = 2, B* = 2.3 T

W = 200 kW, w = 40 MW/m³

$L = 0.7 \text{ m}$

Inlet dimension 24.9 mm x 99.8 mm

Exit dimension 72.6 mm x 114.3 mm

Base Design #2--Russian PAMIR-1 (ref. 8)

Solid fuel ($c = 2.22 \times 10^{-33} \text{ (mho/m)(atm)}^{0.5} \text{ (K)}^{-10}$)

$p^0 = 45 \text{ atm}$, $T^0 = 3559 \text{ K}$

$M_1 = 2.14$, $B^* = 4 \text{ T}$

$W = 15 \text{ MW}$, $w = 500 \text{ MW/m}^3$

$L = 1 \text{ m}$

Inlet dimension 160 mm x 140 mm

Exit dimension 160 mm x 220 mm

Base Design #3--Maxwell 30 MW Unit (ref. 11)

JP-4 + O_2 , seeded with Cs

($c = 8.45 \times 10^{-34} \text{ (mho/m)(atm)}^{0.5} \text{ (K)}^{-10}$)

$p^0 = 30 \text{ atm}$, $T^0 = 3530 \text{ K}$

$M_1 = 2.2$, $B^* = 4 \text{ T}$

$W = 30 \text{ MW}$, $w = 200 \text{ MW/m}^3$

$L = 1.3 \text{ m}$

Inlet dimension 200 mm x 200 mm

Exit dimension 450 mm x 450 mm

Base Design #4--AVCO VIKING-1 (ref. 12)

Toluene + O_2 , seeded with Cs

$p^0 = 15 \text{ atm}$, $T^0 = 3100 \text{ K}$

$M_1 = 2.2$, $B^* = 2.8 \text{ T}$

$W = 2 \text{ MW}$, $w = 47 \text{ MW/m}^3$

$L = 1.75 \text{ m}$

Inlet dimension 50 mm x 150 mm

Exit dimension 166 mm x 249 mm

In the calculation, the scaled unit always uses toluene + O_2 ,
seeded with Cs, with $c = 1.68 \times 10^{-33} (\text{mho/m})(\text{atm})^{0.5}(\text{K})^{-10}$.

TABLE IV-1
DESIGN VIA SCALING LAWS

Base Design	Size Factor	T°	4 T	5 T
#1	2x	3100 K	p° = 33.3 atm	44.7 atm
			w = 66.3 MW/m ³	89 MW/m ³
			W = 2.8 MW	3.58 MW
	1.6x	3300 K	51.4 atm	69.3 atm
			106 MW/m ³	143 MW/m ³
			4.25 MW	5.7 MW
#2	1x	3100 K	40.1 atm	
			114 MW/m ³	
			2.35 MW	
#3	0.8x	3100 K		57.2 atm
				157 MW/m ³
				2.6 MW
#2	1x	3100 K	10.5 atm	14.1 atm
			105 MW/m ³	142 MW/m ³
			3.15 MW	4.25 MW
#3	0.8x	3100 K	10.8 atm	14.6 atm
			78 MW/m ³	105 MW/m ³
			6 MW	8 MW

TABLE IV-1

(CONTINUED)

#4	1x	3100 K	24 atm 75.5 MW/m ³ 3.2 MW	32.5 atm 101 MW/m ³ 4.3 MW
	0.9x	3100 K	22.5 atm 78 MW/m ³ 3.3 MW	30 atm 106 MW/m ³ 4.5 MW
#1	1.8x	3100 K	31 atm 68.7 MW/m ³ 2 MW	44.7 atm 92.6 MW/m ³ 2.7 MW
		3300 K	47.9 atm 110 MW/m ³ 3.21 MW	64.6 atm 148 MW/m ³ 4.3 MW

SECTION V

LOSSES

The losses near the walls and ends, which are totally ignored in the foregoing sections, will eventually be accommodated in the computer simulation that must intercede between the preliminary design and the actual construction. However, a few qualitative remarks about wall effects can still be quoted here.

First of all, since MHD generation is a volume phenomenon, while wall losses are surface phenomena, the percentage loss out of the total power must be proportional to

$$\frac{(\text{Surface of the generator})}{(\text{Volume of the generator})} \propto \frac{1}{L}$$

Thus, a larger unit (scaling up) will suffer less from the wall losses, and will have a larger efficiency. (The same trend is to be expected also regarding the end losses, since the end regions will occupy a smaller percentage for larger units.) But, as a penalty, a larger unit will be more difficult to cool (to keep the wall material at a manageable temperature level) since the (volumetric) Joule heating per unit surface area (available for cooling) will be proportional to L .

Secondly, there is the Reynolds number as a measure of the viscous effect, which is completely ignored in the preliminary design. As a rule of thumb, the scaling should not change the Reynolds number by a factor of more than 3. Obeying this rule, one is usually sure that

the effects will change only quantitatively by a rather slight degree (e.g., the friction coefficient is roughly proportional to the one-fifth power of the reciprocal of the Reynolds number, also the disturbance due to wall roughness becomes less for larger units). But, a drastic change in Reynolds number must always be scrutinized carefully, for fear that some qualitative deviation may evolve; the boundary layer may separate from the wall, for instance.

REFERENCES

1. Smith, J. M. and L. D. Nichols, "Estimates of Optimal Operating Conditions for Hydrogen-Oxygen Cesium-Seeded Magnetohydrodynamic Power Generator," NASA TN D-8374, 1977.
2. Lu, P.-C., Introduction to the Mechanics of Viscous Fluids, McGraw-Hill Book Co., New York, 1977.
3. Lu, P.-C., Fluid Mechanics: A First Course, Iowa State University Press, Ames, Iowa, 1978.
4. Rosa, R. J., Magnetohydrodynamic Energy Conversion, McGraw-Hill Book Co., New York, 1968.
5. Raeder, J. and G. Zankl, "Influence of the Combustor Parameters on the MHD Generator Performance," Workshop on the Performance of Combustion MHD Generators, Munich, June 26-27, 1972.
6. Bender, E., "One-Dimensional Flow in MHD Generators," AIAA J., Vol. 3, pp. 167-169, 1965.
7. Sonju, O. K., Teno, J., Lothrop, J. W. and S. W. Petty, "Experimental Research on a 400 kW High Power Density MHD Generator," AFAPL-TR-71-5, Air Force Aero Propulsion Laboratory, Wright-Patterson Air Force Base, Ohio, May 1971.
8. Maxwell Laboratories, Inc., "High Power Study Final Briefing Presentation Material," for the Air Force Aero Propulsion Laboratory, Wright-Patterson Air Force Base, Ohio, 12 January 1976.
9. Garrison, G. W., Brogan, T. R., J. J. Nolan, et. al., "Development of Design Criteria, Cost Estimates, and Schedules for an MHD High Performance Demonstration Experiment," AEDC-TR-73-115, Arnold Engineering Development Center, Arnold Air Force Station, Tenn., August 1973.
10. Shanklin, R. V., III, Lytle, J. K., Nimmo, R. A., Buechler, L. W. and H. W. Hehn, "KIVA-I Extended Duration MHD Generator Development," AFAPL-TR-75-27, Air Force Aero Propulsion Laboratory, Wright-Patterson Air Force Base, Ohio, June 1975.
11. Sonju, O. K. and J. Teno, "Study of High Power, High Performance Portable MHD Generator Power Supply Systems," AFAPL-TR-76-87, Air Force Aero Propulsion Laboratory, Wright-Patterson Air Force Base, Ohio, August 1976.

AD-A065 650

OHIO STATE UNIV RESEARCH FOUNDATION COLUMBUS
USAF-ASEE (1978) SUMMER FACULTY RESEARCH PROGRAM (WPAFB). VOLUM--ETC(U)
NOV 78 C D BAILEY

F/G 1/3

F44620-76-C-0052

AFOSR-TR-79-0231

NL

UNCLASSIFIED

4 OF 6

AD
A065650



12. Kessler, R., Sonju, O. K., Teno, J., Lontai, L. and D. Meader,
"MHD Power Neneration (Viking Series) with Hydrocarbon Fuels,"
AFAPL-TR-74, Air Force Aero Propulsion Laboratory, Wright-Pat-
terson Air Force Base, Ohio, November 1974.

LIST OF SYMBOLS

A	cross sectional area of (the core of) the generator
a	a constant, see Eq. (15)
B	magnetic field strength
B*	peak magnetic field strength
b	proportionality constant of β vs. p
c	a proportionality constant, see Eq. (18)
c _p	specific heat capacity at constant pressure
E	electric field strength
F()	a function
j	electric current density
K	loading factor
L	length of generator
M	Mach number
m	an index
n	an index
p	pressure
p ^o	combustion chamber pressure
R	gas constant
S	a parameter, see p. 19
T	temperature (Kelvin)
T ^o	combustion temperature (Kelvin)
u	flow velocity
W	electric power output

w	electric power density (per unit volume)
w_{eff}	effective electric power density
x	coordinate in the flow direction
y	a coordinate direction, see Fig. III-1
α	(Hall field)/(Faraday field)
β	Hall parameter
γ	ratio of specific heat capacities
ρ	density
σ	electric conductivity
$()^*$	reference quantity
$()'$	differentiation with respect to x or (x/L)
(\sim)	dimensionless quantity
$()_1$	at the generator inlet
$()_x$	x -component
$()_y$	y -component
$()_1$	of base design
$()_2$	of scaled (model) design

Units:

atm	(standard) atmosphere, a unit of pressure
MW	megawatt
T	tesla (10^4 gauss)

AIR FORCE MATERIALS LABORATORY

Research Associates:

Brian K. Lambert, Texas Tech University

George K. Miner, University of Dayton

Robert E. Young, Wayne State University

TECHNICAL MEMORANDUM AFML/LLM-78-1

ECONOMIC MODELING AND COMPUTER-AIDED
PROCESS PLANNING FOR SHEET METAL OPERATIONS

Brian K. Lambert
Texas Tech University
Lubbock, TX 79409

Prepared for: USAF-ASEE Summer Faculty Research Program (WPAFB)

FOREWORD

This report was prepared by Dr. Brian K. Lambert while serving as a Faculty Fellow in the USAF-ASEE Summer Faculty Research Program (WPAFB) from June to August, 1978. Dr. Harold L. Gegel, AFML/LLM, served as Research Colleague. The program was administered by Dr. Cecil Bailey, Aeronautical and Astronautical Engineering, The Ohio State University, Columbus, Ohio.

The author wishes to express his appreciation to the Metals and Ceramics Division, Air Force Materials Laboratory for their support and hospitality during the program. In particular, he would like to express appreciation for the assistance of his Research Colleague, Dr. Harold Gegel. The financial support of the Air Force Office of Scientific Research through the USAF-ASEE Summer Program is gratefully acknowledged.

This report has been reviewed and is approved.

Harold L. Gegel
HAROLD L. GEGEL

Senior Scientist
Processing and High Temperature Materials Branch
Metals and Ceramics Division
Air Force Materials Laboratory

DISTRIBUTION:

OSU (5 cys) (Dr. C.D. Bailey)
AFOSR/NE (Dr. A.H. Rosenstein)
AFML/CC (Dr. F.N. Kelley)
AFML/CA (Dr. H.M. Burte)
AFML/LL (Dr. N.M. Tallan)
AFML/LLM (Dr. H.L. Gegel)
AFML/LLM (Mr. A.M. Adair)
AFML/LT (Mr. D.E. Wisnosky)
AFML/XR (Mr. B. Emerick)
AFML/LTM (Mr. H.A. Johnson)
AFML/LT (Mr. J.J. Mattice)

ECONOMIC MODELING AND COMPUTER-AIDED
PROCESS PLANNING FOR SHEET METAL OPERATIONS

Brian K. Lambert
Texas Tech University

ABSTRACT

A program for developing an economic model and a computer-aided process planning system for sheet metal operations is outlined as a part of the processing science research effort of the Air Force. Such a program is intended to enhance the integration of design, planning, and manufacturing functions into a cohesive system. An overall approach for developing a system that will be usable by designers for economic evaluation of material-configuration alternatives and by process planners for determining cost effective process operations and sequences is described. The economic model concept is based on establishing the relationships between part features and the cost components of setup, processing, handling and tooling. The computer aided process planning system will be based on the use of decision tables for determining the technological desirability of alternative processing methods at each stage of production. Such alternative process plans will act as input to the economic model for identifying cost optimum operations and sequences for aircraft sheet metal parts.

I INTRODUCTION AND BACKGROUND

A complex problem currently being encountered by many industries is the enhancement of the integration of design, planning, and manufacturing functions. One approach to this problem is the utilization of the design/manufacturing system concept in which a computerized information flow and decision making process exists throughout the factory.

The use of computers as an aid in solving specific design and manufacturing problems has been in existence for some time; however, relatively little effort has been made to integrate the various individual applications into a cohesive system. An extensive plan to achieve such a unified system is the Air Force Integrated Computer Aided Manufacturing Program (ICAM). The ICAM program is not intended to be a mechanism for additional substitution of computers for man-controlled functions, but rather may be described as follows:

"ICAM is a system that uses computers to organize every step of manufacturing- from parts design to physical location of machine tools to shipping- in the most economical and efficient mode." (1)

Although the program has numerous objectives, a conceptual goal of considerable importance to managers and engineers is:

"A generative planning capability whereby a part designer could not only optimally design a part, but at the same time subject this part to a performance evaluation, and plan for the most economical fabrication of the part within the constraints of schedule, availability of raw materials and variability of materials or processes. Further, it could be envisioned that the fabrication test could be performed immediately and the part production automatically introduced into the overall manufacturing plan." (2)

In establishing the basic approach to the ICAM program, the concept of a manufacturing "wedge" was introduced. A wedge is defined as, "A collection of subsystems which on the shop floor are specific to a particular process - but in other levels of manufacturing - may be only a part of general support, management and control subsystems"(3). The manufacturing technology selected for the initial wedge concept treatment was sheet metal processing. Several factors led to the selection of this particular group of operations, one reason being that, unlike machining, the aerospace industry has invested only to a moderate degree in advanced techniques for sheet metal fabrication and improvement should rapidly result in productivity increases and reduced costs (4).

As indicated by the previous statements pertaining to the basic concepts and goals of the ICAM program, an important and necessary aspect of the program is the development of economic models for various manufacturing technologies - the first being for sheet metal fabrication. Such models will enable the performance of material - configuration - cost tradeoff studies during the early portion of the design phase as well as the economic evaluation of alternative processing methods.

At the present level of manufacturing planning technology, process planners, in most cases, rely on their own experience to select the processing operations and sequence to be utilized. Similarly, designers must generally rely on experience to evaluate the economic consequences of alternative part designs. Thus, both the

design and process planning functions are "experienced based" rather than "knowledge based". This situation does not allow an individual designer or planner ready access to a firm and broad-based technology data base which could form the foundation for systematic evaluation of alternatives. Consequently, a need exists for a planning capability for evaluating the economic aspects of design and processing alternatives. In fact, this need is apparent in all phases of the design - manufacturing system and is described as follows:

"As evidenced by the description of stages that follows, the need for the concept of a component/subassembly/assembly evaluation of economic alternatives is present until final production: (1) Concept/design stage: which of the manufacturing processes that are capable of producing components to the desired specification, are low cost and can be executed in the available or potentially available plant facilities? (2) Process/material development stage: what are the cost effective manufacturing processes and materials that can be developed to meet concept and design requirements? (3) Cost estimation/value analysis/pre-planning stage: of the available or potentially available processes, which sequence will yield economical manufacture of a given batch or series size of a given component/subassembly/assembly? (4) Detailed process planning stage: which sequence of operations and which operating conditions for each operation should be chosen so as to minimize the total cost of manufacturing a given batch or series size of a given component/subassembly/assembly? (5) Production scheduling and control stage: how should production of a given variety of components through given production facilities be scheduled so as to produce acceptable parts on time and at a minimum cost?"(5)

PURPOSE

The purpose of this study is to structure an overall program which will lead to the development of an economic model and computer-aided process planning system for sheet metal operations. The model and planning system will be usable by designers for economic evaluation of material-configuration alternatives and by process planners for determining cost effective process operations and sequences. In addition, the economic model and process planning system will be constructed in such a manner to provide for integration into the overall ICAM program as well as the Air Force processing science research effort.

REVIEW OF PROCESS ECONOMIC MODELING

The need and potential applications of economic models of manufacturing processes were described in the previous section. The following discussion presents a brief summary of the current status of economic modeling at the process level.

Interest in the economic analysis of manufacturing processes has been concentrated primarily on machining operations, beginning with F.W. Taylor's work in the early 1900's. During the last 20 years researchers have investigated numerous aspects of the machining economics problem with the major emphasis being on the determination of the levels of the machining parameters (speed, feed, depth of cut, etc.) to achieve some specified criterion such as minimum production cost, maximum production rate, or maximum profit rate. Investigations have also been conducted concerning statistical and probabilistic models, multiple station and multiple tool machining systems (5). Recently, various optimizing techniques such as response surface methodology (6,7), geometric programming (8), chance constrained

programming (9), convex mathematical programming (10) and others have been applied to the problem.

The utilization of computers as an aid to analyzing and planning machining operations has also received considerable attention in the last few years (11, 12, 13, 14, 15).

Research in the area of machining has been of interest not only because of the economic significance of the process but also because of the process parameter dependent nature of machining. That is, the cost of machining operations can be directly related to the process parameters and thus determination of the levels of the machining variables to achieve minimum cost or maximum rate conditions is mathematically feasible. This situation is not necessarily the case for certain other classes of metal working processes. In sheet metal operations, for example, the actual process variables such as punch speed have relatively little effect on overall cost when compared to setup, handling, and tooling costs.

With respect to the development of basic principles and methodologies pertaining to process economic analysis, relatively little work has been done. Colding has discussed some general aspects of cost modeling (16,17) and classifies manufacturing costs into six categories: (1) methods, planning and scheduling, (2) design and manufacture of jigs and fixtures, (3) "pure" shop costs, (4) material, (5) inventory, and (6) extra costs.

A literature review revealed that little effort has been directed towards developing an economic model for sheet metal operations. As previously mentioned, unlike machining, sheet metal operations are not highly process parameter dependent. That is, no process variables comparable to the machining variables of cutting speed and feed have a dominant effect on total cost. Sheet metal operations appear to be more dependent on part material, shape, and size which, in turn, influence setup, handling and tooling costs and thus total cost.

One case study has been performed in an attempt to establish a computer aided work planning system for simple sheet components (18). By inputting a part code, the program reportedly outputs blank size, material costs, an operations sequence, setting and cycle times, and machining cost. Only some 400 parts of simple configurations were considered.

II OVERVIEW OF PROPOSED SYSTEM

An initial step in the development of an economic model for sheet metal processing is the identification of the materials currently being used, range of part sizes, lot sizes, and types of processes utilized. A preliminary analysis has been done by the Boeing Company under the CAM Architecture - Task III (Sheet Metal Fabrication Technology) effort. The major results of this study are summarized as follows (19):

1. Material distribution: Of the 8.5 million sheet metal parts on file, the material usage distribution was: aluminum 87.7%, stainless steel 6.9%, carbon steel 3.4%, titanium 1.6%, nickel base and other alloys .4%. Examination of survey data from other companies indicated a similar distribution.
2. Part size distribution: 90 percent of the parts considered are produced from sheet less than .10 inch thick in blank sizes not exceeding 8 X 25 inches.
3. Lot size distribution: 90 percent of the lot sizes are 160 parts or less with most lots averaging 20 parts or less.
4. Process distribution: Figure 1 shows the breakdown of trimming and forming operations. The data indicates that 57 percent of the parts are trimmed by sequential shearing and blanking operations and 23 percent are sheared only. Thus, these two processes account for 80 percent of the trimming operations. For forming, 57.4 percent of the parts are brakeformed and 27.8 percent are hydroformed; these two processes constitute 75.2 percent of all the forming operations. Results of an industry survey indicate substantial agreement in the relative utilization of forming processes.

This study did not provide information concerning additional processing steps such as cleaning, surface treatment, deburring, heat treatment, etc., nor was information included concerning the costs of the various operations.

5. Part shape distribution: Analysis of part shapes indicates the following: one-bend parts 29.3%, two or more parallel bends in one direction 22%, two or more non-parallel bends in one direction 10.5%, two or more parallel bends in two directions 8.9%, two or more non-parallel bends in two directions 15.5%, and curved bend line 13.2%.

Comments in the Boeing report further indicate the need for an economic model of sheet metal operations: "Since several different processes contribute to the total time, interface to a total-part-cost computer model is necessary. In this model, standard cost elements for a sequence of processes can be added to give an estimate of the total part cost"(19). The report also states, "Generally, there is no means for selecting processes based on economic considerations. Without such an analytical capability, engineering will not be able to effectively conduct component cost trade studies based on formability data during the design state, nor will it be possible to most effectively implement a generative planning capability"(19).

To perform an evaluation of alternative part designs and processing sequences, the basic system structure shown in Figure 2 could be utilized. Based on design requirements such as stiffness, strength, weight, etc., the designer provides an initial attempt at specifying the material type, size, shape and other relevant characteristics which will meet the functional requirements of the part. This information provides the input to the next portion of the system.

If a group technology system is available, then the alternative part design under consideration can be classified and coded and the process plan for the family

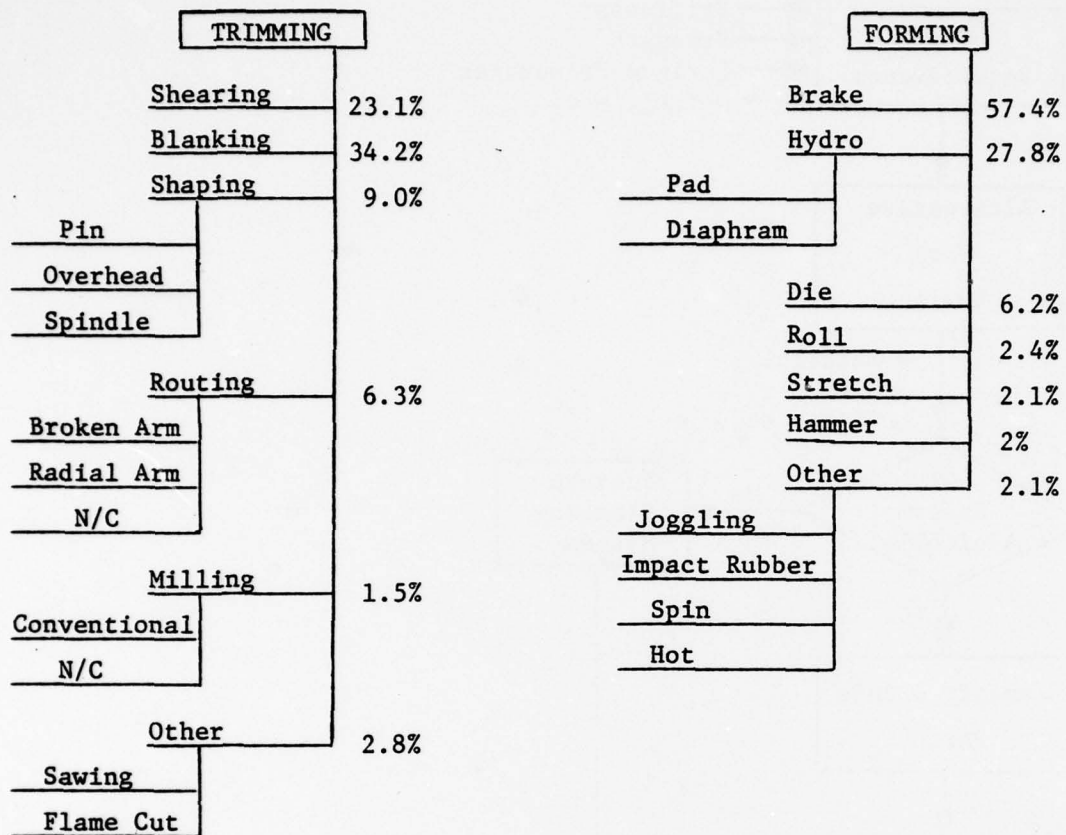
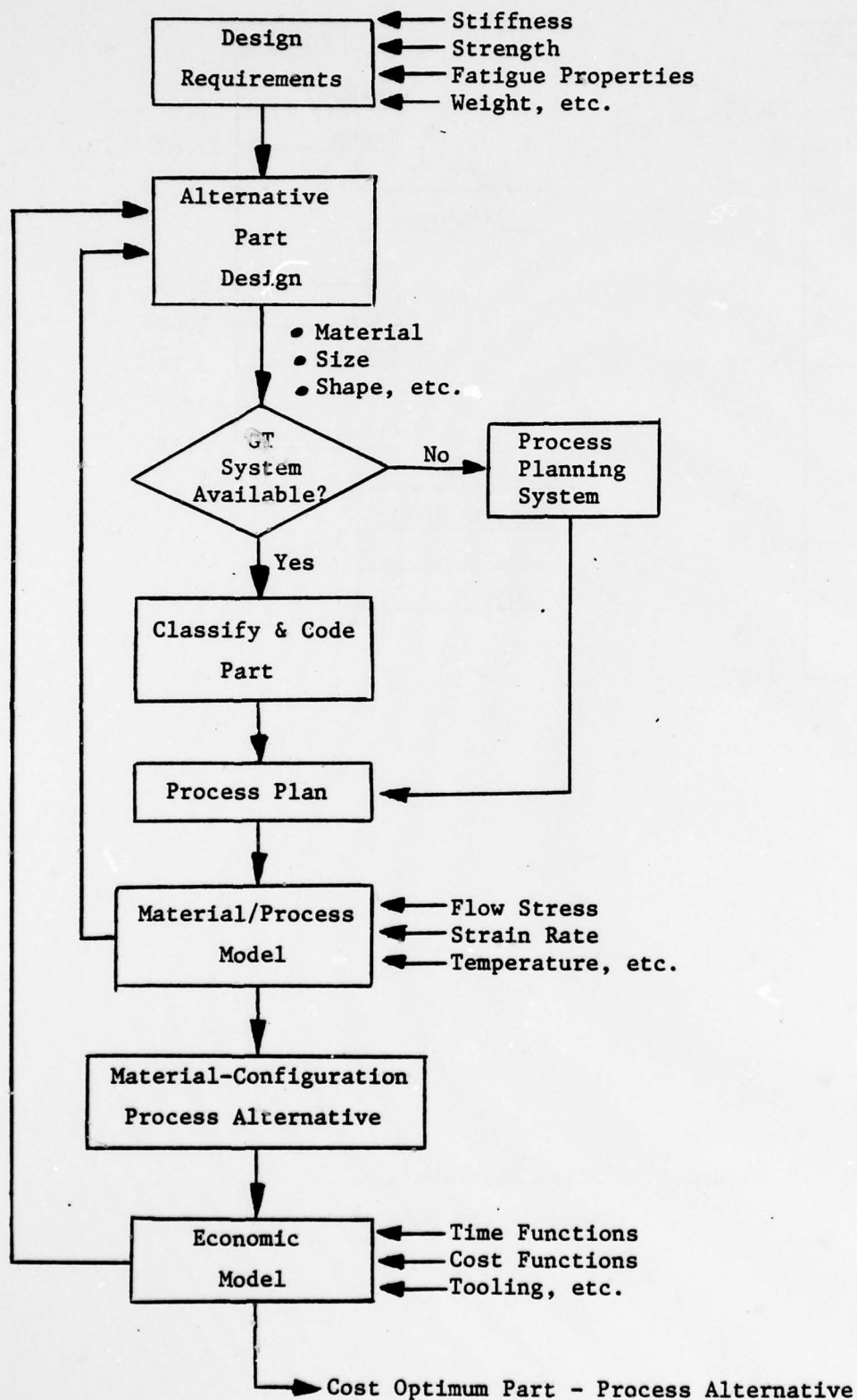


Figure 1 Process Distribution
(Adapted from Reference 19)

Figure 2 Basic System Structure



to which the part belongs can be generated. If a group technology system including process plans for various families is not available, then a process planning system must be utilized. Such a system will be discussed in a later section. At any rate, the alternative part design and process plan act as an input to the material/process model portion of the overall system.

The development of the material and process models is currently underway at Battelle Laboratories (Columbus) and is intended to provide analytical models describing the material behavior under various conditions of temperature, strain, strain rate, etc. The model for material behavior will describe the limits to which the material can be deformed and the process model will describe the local states of stresses and strains in the material during forming (20). Essentially then, the material model answers the question of what is the material capable of doing and the process model answers the question of what is the process asking the material to do. The net result of the application of these models is the determination of whether or not the material and process can achieve the desired part characteristics. At this stage, revisions to the part design may be necessary with outputs from the material/process model acting as feedback to the design function. Once this cycle has been completed, a feasible material - configuration - process alternative will have been identified. This information can then be transmitted to the economic model and the cost of the alternative under consideration can be determined. The entire procedure can then be repeated for other material - configuration alternatives and the most cost - effective combination selected.

III ECONOMIC MODEL FORMULATION

In constructing an economic model for sheet metal operations it is first necessary to identify all the processing steps required to produce a given part. This includes not only the trimming and forming operations (discussed in the Boeing Task III report) but also other processing stages such as deburring, straightening, cleaning, heat treatment, hand operations, etc. Methods for determining the processing steps required for a given material - configuration will be discussed in the section on Process Planning Methods. In addition to specifying the operations to be performed, the types of costs to be considered must be determined and the relationships between these various costs and some set of part characteristics must be established. Ultimately, predictive equations for manufacturing costs as a function of part characteristics must be developed.

On a macro basis, the total cost for producing a manufactured product would include direct materials and labor, factory overhead, commercial expenses, and selling cost. These major cost components are described in Table 1, and examples of the various elements which make up these costs are illustrated. This study, however, is concerned with an economic model at the individual part or process level (what might be termed a micro-economic model) and will concentrate on material costs, labor costs, and manufacturing overhead. The interpretation of these costs for the purposes of this study are discussed in the following paragraphs.

The direct material cost refers to the cost of raw material or semi-finished material directly traceable to an operation or part. A general expression for the direct material cost can be formulated as:

$$C_m = W(L_1 + L_2 + L_3)P_m - R$$

where C_m = material cost/unit

W = weight of a unit or in other compatible dimensions relating to price

L_1 = scrap loss

L_2 = waste loss

L_3 = shrinkage loss

P_m = price per pound, length, volume, or area

R = anticipated salvage value of material

Here, scrap loss refers to rejection of defects consisting of unusable raw material including improper material layout, fracture of formed parts, etc. Waste loss is illustrated by that material remaining after parts have been blanked out in stamping operations. Shrinkage losses may occur in certain processes or materials such as wood and plastics.

The general expression for direct material cost can be modified for specific processes. For example, in a sheet metal stamping operation (assuming no salvage):

$$C_m = WLtP_m(1+e)$$

where W = stock width

L = blank length

t = stock thickness

e = expected losses

TABLE 1 - TOTAL COST COMPONENTS

Major Cost Component	Sub-Components	Typical Elements
Prime Cost	{ Direct Materials	- Sheet Metal, Forging
	{ Direct Labor	- Setup, Processing
Factory Overhead	{ Indirect Materials	- Supplies, Lubricants
	{ Indirect Labor	- Supervision, Clerks
	{ Fixed & Miscellaneous	- Rent, Utilities
Commercial Expenses	{ Distribution	- Advertising, Samples
	{ Administrative	- Legal, Auditing
Selling Cost		- Salaries, Commissions

The stock width is determined by the width of the part, W_p , plus the allowable margins on each side of the blank. It is possible (for a given material) to express the allowable margin as a function of stock thickness, say At . Then:

$$W = W_p + 2At$$

The length required is the part length, L_p , plus the required distance between parts, again expressed as a function of thickness, Bt . Thus:

$$L = L_p + Bt$$

The direct material cost then becomes:

$$C_m = (W_p + 2At)(L_p + Bt)tP_m(1+e)$$

Indirect materials are those that are necessary for operation in some manner but are not traceable to a specific operation or component. Cutting fluids, lubricating oils, clerical supplies, etc. are examples of such materials. Because the cost of indirect materials is difficult to assess on the process or product level, the cost is usually charged to a given operation by means of overhead distribution. Indirect labor costs are handled in generally the same manner as indirect materials cost.

Direct labor may comprise one of the most important items of the manufacturing cost and is defined as the cost of actually producing goods, or the labor cost directly traceable to a given part or operation. The direct labor cost directly influences the setup cost, processing cost and handling cost of an operation. One means of expressing these costs is as follows:

$$C_{pi} = R_{pi}T_{pi}$$

$$C_{si} = \frac{R_{si}T_{si}}{Q}$$

$$C_{hi} = \frac{R_{hi}T_{hi}}{H}$$

where C_{pi} = processing cost for operation i
 C_{si} = setup cost for operation i
 C_{hi} = handling cost between stages
 R_{pi} = labor rate for operation i
 R_{si} = setup rate for operation i
 R_{hi} = labor rate for handling
 T_{pi} = processing or run time for operation i
 T_{si} = Setup time for operation i
 T_{hi} = handling time between stages
 H = number of parts per unit handling load
 Q = quantity or lot size

For convenience, handling time at a processing stage or machine is included in the run time, T_{pi} , and the setup time, T_{si} , includes setup for a given operation as well as tear down to clear the machine for the next operation.

The manufacturing overhead cost (similar terms include: overhead expense, burden, indirect expense, factory expense, etc.) can be defined as that portion of the cost which is not clearly associated with particular operations and must be prorated among all the cost units on some arbitrary basis. Essentially, manufacturing overhead includes all production costs except direct materials and labor. Numerous methods exist to distribute manufacturing overhead to jobs. These costs may be applied on the basis of: (1) direct labor hours or cost, (2) direct material cost, (3) direct labor cost plus direct material cost, (4) unit of product, (5) machine hours, and others. Each method has certain advantages and disadvantages and the method utilized varies from company to company and, in some cases, within the various sections in a company.

The determination of the tooling costs for a given operation creates some special problems depending on the type of tooling required and the magnitude of the costs involved. Considering its importance to production, determination of tooling cost calls for special attention and is an exception to the practice of ignoring indirect costs. One method for expressing tooling is to consider it to be composed of two parts: A usage or depreciation cost and a reconditioning or rework cost. The usage cost, C_u , may be expressed as:

$$C_u = \frac{C_o}{1 + N_1 N_2}$$

where C_o = initial cost of tool

N_1 = number of parts made before the tool must be reworked

N_2 = number of times the tool can be reworked before discarding

The reconditioning cost, C_r , may be written as:

$$C_r = \frac{R_r T_r}{N_1}$$

where R_r = labor and overhead for tool rework

T_r = time to rework a tool

The tooling cost for a given operation is the sum of these two components:

$$C_t = \frac{C_o}{1 + N_1 N_2} + \frac{R_r T_r}{N_1}$$

Note that the same type of formulation can be used for inspection and gaging operations.

Having described the basic costs involved on the micro level, it is now possible to write a general expression for the manufacturing cost for a given part. The situation is schematically shown in Figure 3 for a material undergoing n processing steps to produce a given part, and the total cost equation (without overhead) is:

$$TC = C_m + \sum_{i=1}^n (C_{si} + C_{pi} + C_{ti}) + \sum_{i=0}^n C_{hi}$$

$$TC = C_m + \sum_{i=1}^n \left(\frac{R_{si} T_{si}}{Q} + R_{pi} T_{pi} + \frac{C_{oi}}{1 + N_{1i} N_{2i}} + \frac{R_{ri} T_{ri}}{N_{1i}} \right) + \sum_{i=0}^n \frac{R_{hi} T_{hi}}{H_i}$$

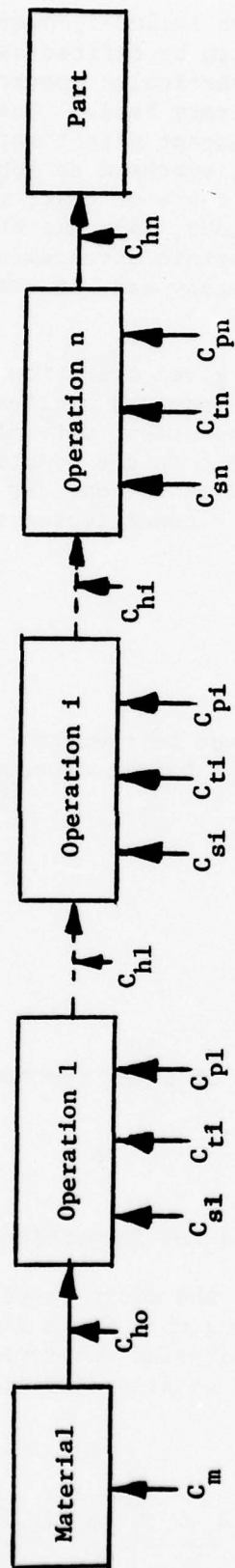


Figure 3 Processing Cost Components

The manufacturing overhead cost can be added to this expression depending on the method that is used. For example, if overhead is based on total direct labor hours, then the overhead cost allocated to unit cost is:

$$C_{OH} = R_{OH} \sum_{i=1}^n (T_{si} + T_{pi} + T_{hi} + T_{ri})$$

where R_{OH} = overhead rate as a decimal.

DETERMINING MODEL PARAMETERS

In order to utilize the proposed economic model for evaluating alternative part designs or processing methods, relationships between part characteristics and cost or time values must be established. The economic model parameters which must be estimated include:

- 1) setup time
- 2) run or processing time
- 3) handling time
- 4) initial tooling cost
- 5) tool life
- 6) tool reconditioning time
- 7) material requirements including scrap loss

These parameters must be related to certain part characteristics which might include the following:

- 1) material type
- 2) length, width, thickness
- 3) tolerances
- 4) number of bends, radii, direction
- 5) part shape
- 6) holes
- 7) joggles, beads
- 8) surface condition requirements
- 9) part identification requirements

This list is by no means complete, and other part characteristics which influence manufacturing cost will be identified upon examination of information gathered from industry.

It is anticipated that multiple regression analysis will be used to develop the relationships between part characteristics and the economic parameters. Equations which will predict the values of the various economic parameters as a function of part characteristics will be developed. Application of such a method will also allow the determination of confidence intervals and ranges on manufacturing time and cost for various alternatives. In addition to showing the main effects of part characteristics on costs, any interacting effects among the variables can be identified.

IV PROCESS PLANNING SYSTEM

As mentioned earlier, utilization of an economic model for sheet metal operations requires an input which specifies a process plan, i.e., the operations and sequence necessary to fabricate the part. Consequently, it is logical to consider developing a computer aided process planning system in conjunction with the development of an economic model. The advantages of automated process planning systems are numerous and the feasibility of such systems has been well established in the area of machining.

Two basic approaches for constructing an automated process planning system exist. One method is to establish a file of coded parts and their associated process plans. A new part can then be coded, the file searched for an identical or similar part, the process plan retrieved and modified as necessary. The second method involves generating a process plan from part characteristics in an analytical manner. This method requires determining the connection between part features and processing operations. Actually, a combination of both methods would appear to be the most desirable.

The overall problem may be viewed as a sequential decision process where the decision made at one stage may have a significant influence on subsequent decisions. The selection of a particular metal may dictate certain processing steps; for example, deburring of brake form blanks prior to forming may be necessary in order to prevent end cracks or end displacement depending on the metal selected. Part size, shape, tolerances, finish requirements, etc., will influence the alternatives available at each processing step.

Just as the economic model will require relationships between part characteristics and various cost components, the process planning system will need relationships between part features and processing steps. In fact, the data base requirements for the economic modeling effort and the process planning effort are highly similar, which enhances the idea of concurrent development. By inputting certain information about the part, alternative methods for processing at each stage can be generated. These process plans can then be evaluated by using the economic model and the lowest cost method identified.

One approach to be considered for determining feasible operations at each stage of processing would be the use of decision tables. These tables would indicate which processing methods are potential candidates based on certain relevant part characteristics. An example of a decision table for forming method versus material is presented in Table 2. The numbers in the matrix could be indicators of the relative technological desirability of the forming method-material combination. The indicators could reflect such factors as tolerance control, effect on material properties, process difficulties, etc. Cost considerations would not be included since the cost of the various alternatives will be determined using the economic model.

After examining the material versus forming decision table, the computer would continue to the next part parameter (shape, size, etc) versus forming method and so on until all relevant part features have been examined. At this point, a listing of all feasible methods for the processing step of forming along with a technological desirability rating of the methods would be available. This procedure would continue for each processing step and alternative process plans would be generated. These plans would then act as input to the economic model for evaluation. Thus, both a technological desirability and cost evaluation would be conducted.

TABLE 2
DECISION TABLE: MATERIAL-FORMING METHOD

		Forming Method					
		F_1	F_2	F_3	F_4	F_5	$\dots F_n$
		Brake- form	Hydro- form	Die form	Roll form	Stretch form	
Material	M_1	Aluminum	5	3	2	2	1
	M_2	Stainless Steel	5	3	2	1	1
	M_3	Carbon Steel	5	2	1	1	1
	M_4	Titanium	5	2	1	0	1
	M_5	Ni Base	5	2	1	0	1
	\vdots	\vdots					
	M_m						

V PROGRAM PLAN

The previous sections contained discussions indicating the need and potential usefulness of an economic model and computer aided process planning system for sheet metal operations. Certain conceptual approaches to the problem were also described. The following paragraphs contain an outline of the various tasks which will be necessary to develop an automated process planning system and an economic model.

1. Review of previous research. A continuation of the literature review should be undertaken in three major areas:

- (a) process economic modeling - general principles and methods and cases dealing specifically with sheet metal operations
- (b) computer aided process planning systems
- (c) technical information pertaining to alternative processing methods for sheet metal parts.

2. Establish information sources. In order to establish a data base for developing the economic model and planning system, cooperation with industry will be necessary. Potential contacts should include:

- (a) aircraft industry
- (b) other industries which might have useful information
- (c) trade associations and professional societies
- (d) equipment and tool manufacturers

3. Establish data base. The information required to develop the proposed model and planning system should include:

- (a) process plans for typical sheet metal parts showing part configuration and processing steps from raw material stores to finished part.
- (b) manufacturing time and cost data for setup time, processing time, tooling requirements, and handling time.
- (c) raw material requirements and cost
- (d) tool life and tool reconditioning costs

4. Analysis of process plans. The processing plans for various parts will be analyzed to:

- (a) identify processing sequences for different materials and configurations
- (b) identify alternative processing methods at each production stage
- (c) identify equipment types and uses.

5. Part characteristic analysis. This task will involve determining the part features which influence:

- (a) processing sequence
- (b) setup time and cost
- (c) processing time and cost
- (d) handling time and cost
- (e) tooling requirements and costs

6. Develop predictive equations for economic model. Information derived from tasks (4) and (5) will be used to generate predictive equations for:

- (a) time components
- (b) cost components

7. Construct computer aided economic model. The predictive equations developed in task (6) will form the basis for the computerized economic model. Construction of the computer model will include:

- (a) overall program structure development
- (b) specification of necessary inputs
- (c) development of cost files
- (d) specification of output formats

8. Structure process planning system. This task is concerned with structuring a computer aided processing planning system based on the information gained in tasks (4) and (5) and will include:

- (a) development of overall program structure
- (b) identification of factors for process decision tables
- (c) determination of technological desirability factors and ratings of processing alternatives.
- (d) specification of input requirements in the form of part features
- (e) structuring of output requirements and formats

9. Integration with existing and proposed ICAM systems. Throughout the the proposed program, close coordination with the Air Force ICAM effort will be necessary. Specifically, the economic model and process planning system will be integrated with:

- (a) process and material models
- (b) group technology systems
- (c) other relevant ICAM programs

10. Testing and verification. This task will consist of testing and verification of the economic model and process planning system. Existing parts will be selected and the results of the model and planning system will be compared to industry results which have been generated by other methods.

11. Develop utilization plan. This effort will be aimed at devising methods for implementing the system in industry through short courses, seminars, published reports and papers, etc.

12. Program Assessment. Feedback from system users will be evaluated and any recommendations or necessary modifications will be incorporated.

REFERENCES

- (1) Zimmerman, M.D., "ICAM - Revolution in Manufacturing", Machine Design, May 26, 1977.
- (2) Wisnosky, Dennis E., "Planning for Integrated Computer Aided Manufacturing", SME Manufacturing Management Productivity Opportunities Conference, Dearborn, MI, May 25, 1976.
- (3) Wisnosky, Dennis E., Harris, W.A., and Shunk, D.L., "An Overview of the Air Force Program for Integrated Computer Aided Manufacturing (ICAM)", SME Technical Paper MS77-254, 1977.
- (4) Wisnosky, Dennis E., ICAM Program Prospectus, Air Force Materials Laboratory, Wright-Patterson AFB, Dec. 1, 1977.
- (5) Tipnis, V.A., et.al., Mathematical Modeling of Materials Removal Processes for Improved Process Design, Planning, Optimization and Control. Technical Report AFML-TR-77-154, Metcut Research Associates, Inc., Cincinnati, OH, Sept., 1977.
- (6) Wu, S.M., "Tool Life Testing by Response Surface Methodology, Parts I and II", Journal of Engineering for Industry, Trans. ASME, Series B, Vol 86, 1964.
- (7) Lambert, B.K., and Taraman, K., "Development and Utilization of a Mathematical Model of a Turning Operation", International Journal of Production Research, Vol II, 1973.
- (8) Walvekar, A.G. and Lambert, B.K., "An Application of Geometric Programming to Machining Variable Selection", International Journal of Production Research, Vol 8, 1970.
- (9) Iwata, K., et.al., "A Probabilistic Approach to the Determination of the Optimum Cutting Conditions", Journal of Engineering for Industry, Trans. ASME, Series B, Vol 94, 1972.
- (10) Draghici, G., and Paltinea, C., "Calculation by Convex Mathematical Programming of the Optimum Cutting Condition When Cylindrical Milling", International Journal of Machine Tool Design and Research, Vol 14, 1974.
- (11) Kronenberg, M., "Computerized Determination and Analysis of Cost and Production Rates for Machining Operations", Journal of Engineering for Industry, Trans. ASME, Series B, Vol 91, 1969.
- (12) Crookall, J.R. and Venkataramani, N., "Computer Optimization of Multipass Turning", International Journal of Production Research, Vol 9, 1971.
- (13) Bjorke, O., "Cost Evaluation of Design Features Using AUTOPROS", Manufacturing Systems (CIRP), 1974.
- (14) Barash, M.M., and Berra, P.B., "Automatic Planning of Optimal Metal Cutting Operations and Its Effect on Machine Tool Design", Journal of Engineering for Industry, Trans. ASME, Series B, Vol 93, 1971.

(15) El Gomayel, J. and Abou-Zeid, M.R., "Piece Part Coding and the Optimization of Process Planning", Proceedings, 3rd NAMRAC Conference, Dearborn, MI; SME, 1975.

(16) Colding, B., "A Cost Model and a Performance Index for a Manufacturing System", Annals of the CIRP, 24, 1975.

(17) Colding, B., "The Total Cost Relationship of the Integrated Manufacturing Systems", Annals of the CIRP, Vol 13, No. 2, 1974.

(18) Van Hasselt, R., and Oudolf, W.J., "Computer Aided Work Planned for Simple Sheet Components", Annals of the CIRP, Vol 22, 1973.

(19) Lowery, P.A., Computer Aided Manufacturing Architecture - Task III, Sheet Metal Fabrication Technology, Boeing Military Airplane Development, Technical Report AFML-TR-77-216, Nov. 1977.

(20) Nagpal, V., et.al., Mathematical Modeling of Sheet Metal Formability Indices and Sheet Metal Forming Processes, Interim Technical Report, Battelle Columbus Laboratories, March, 1978.

FINAL REPORT

1978 USAF-ASEE SUMMER FACULTY PROGRAM (WPAFB)

25 August 1978

Participant:

**Dr. George K. Miner
Department of Physics, University of Dayton**

Research Colleague:

**Dr. Patrick M. Hemenger
AFML/LPO**

ABSTRACT

**FEASIBILITY STUDY ON THE USE OF THE EPR
TECHNIQUE ON DETECTOR GRADE SILICON**

The feasibility of employing the electron paramagnetic resonance (EPR) technique in the study of detector grade silicon has been investigated. It was found that data can be collected using an x-band spectrometer in the absorption mode at room temperature and at liquid nitrogen temperature. Also additional experimental directions have been identified involving K-band EPR and the dispersion mode. Other opportunities exist in the use of electron nuclear double resonance (ENDOR). Several dopants and impurities were examined, and future investigations were planned. Interaction between the Participant and the Research Colleague will continue.

This is a report of an effort in the 1978 USAF-ASEE Summer Faculty Program at Wright-Patterson Air Force Base. The Research Colleague and the Laboratory to which the Participant was assigned have as an objective the development of silicon based infrared detector materials. This work requires knowledge and control of various dopants and impurities. The Participant is experienced in electron paramagnetic resonance (EPR), which is a tool with potential for monitoring such dopants and impurities, but one which was not used previously at AFML. The ten-week activity described in this report was directed at evaluating the feasibility of using this technique to work on this problem.

Silicon is the focus of the detector materials program. The anticipated requirement for large detector arrays makes it desirable to apply the extensive technology of semiconductor devices to the infrared detector effort. The general objective for this work is to develop detectors for the 1 to 25 micrometer range. Pure silicon is limited in its response to wavelengths of 1.1 micrometers or less. In order to extend this range, impurity atoms must be incorporated into the silicon. Current dopants of interest include indium doped silicon for response in the range 3 - 5 micrometers and gallium doped silicon for the 8 - 14 micrometer range. These extrinsic detectors may be used in a monolithic integrated circuit technology.¹ The requirement that the detector array have a uniform ability to detect and respond from element to element dictates impurity uniformity, and the precise control of defects. Dopant concentrations range from 10^{16} to 10^{17} per cubic centimeter, and impurity concentrations are as low as 10^{12} per cubic centimeter. This requires careful materials characterization.

The doping of silicon is achieved by introduction during growth, by thermally-controlled diffusion, or by ion implantation. The electrical properties, the time response of the device, its operating temperature, and its optical sensitivity are affected by the presence of impurities. To optimize these characteristics of the detector, careful measurements of optical and electrical properties must be made as a function of temperature. Optical properties of interest include spectral absorption, photoluminescence, and photoconductivity. The electrical transport properties, primarily resistivity and Hall effect, are the responsibility of the Research Colleague.² With these measurements various detector materials are characterized, including the influence of processing methods such as thermal treatment, ion implantation, or transmutation doping.

The performance of silicon infrared detectors is dependent on starting material quality. The degree of purity of the crystal depends on the method of growth. This present work included samples grown by the Czochralski method in quartz crucibles and by vacuum multiple float zoning. Some examination³ of the growing processes has occurred. The float zoning method in general is more attractive because of lower impurity concentrations.

Undesired and often unknown defects have a negative effect on the properties of detectors. These often result from crystal growth or detector fabrication. Heat treatment during processing of the detector, or annealing of ion-implanted or neutron-transmuted material may cause a redistribution of necessary dopants, or may introduce unknown and undesired centers. The EPR technique has been employed in the identification⁴ of a surface defect induced in silicon during the quench after thermal annealing.

EXPERIMENTAL

The laboratory involved does not maintain a magnetic resonance facility. In order to evaluate the EPR technique for use on silicon detector materials, it was necessary for the Participant to make use of the Magnetic Resonance Laboratory in the Physics Department of the University of Dayton, his home institution. That facility now includes two complete electron paramagnetic resonance spectrometers, a nuclear magnetic resonance spectrometer and equipment for ENDOR (electron-nuclear double resonance).

At the beginning of the ten week summer period, the second EPR spectrometer had just been delivered and was not yet operational. Since it has greater sensitivity than the original spectrometer, it was decided that the more recently acquired device should be used for the silicon work since weak signal levels were anticipated. Initial activities included making the spectrometer operational and performing some calibrations.

The device used in this study is a modified form of the Varian Associates model 4500 series EPR spectrometer. EPR spectroscopy is based on the electron's intrinsic angular momentum and its magnetic moment. The ratio of the electron's magnetic moment to its spin value, known as the magnetogyric ratio, is distinct and constant. However, the effects of the local magnetic environment arising from the fields of adjacent nuclei, and from coupling to other angular momenta make the effective magnetogyric ratio unique for each physical environment. When the electron is unpaired, it is possible to detect this ratio by inducing transitions between the electron Zeeman levels, detecting these transitions, and visually displaying them. In this spectrometer

the transitions are induced by electromagnetic radiation in the microwave range. This X-band spectrometer operates at about 9.5 Gigahertz, or about 9.2 Gigahertz with a quartz dewar inserted. The external magnetic field for the Zeeman splitting is provided by a twelve-inch regulated and water-cooled magnet with field scanning provisions. The microwaves are produced by a klystron, are distributed by a microwave bridge and are detected by a crystal detector. The original hybrid tee bridge has been replaced by a more efficient three port circulator arrangement. The klystron is stabilized with automatic frequency control. The spectrometer operates in the absorption mode.

In EPR spectrometers the detector noise varies inversely with frequency, and therefore the higher the field modulation frequency, the better the signal-to-noise ratio. Considering such factors as field modulation, penetration of the cavity walls, sample relaxation times and crystal noise, however, 100 Kiloherztz has resulted as the optimum choice for most applications. This spectrometer has a 100 Kiloherztz crystal controlled oscillator which generates the field modulation frequency, and a high gain amplifier and phase detector for detection of the EPR signal.

The manufacturer states that the sensitivity of the system is " $2 \times 10^{11} \Delta H$ electron spins with a one second integration time, assuming that the sample has negligible dielectric loss." ΔH is the linewidth in gauss. The sensitivity was checked with a standard supplied by the manufacturer. The sample is $3.3 \times 10^{-4}\%$ pitch in KCl. It contains $10^{13} \pm 25\%$ spins. Its linewidth is 1.7 gauss. The observed signal-to-noise ratio for this sample was 13/1.2, or 10.8 ratio. This suggests the possible detection of a minimum of $9.3 \times 10^{11} \pm 25\%$ spins. For this case, a 1.7 gauss line, the specifications give 3.4×10^{11} spins.

The linearity of the magnetic field scan, the Hall field sensing and the X-Y recorder were checked using a known field marker. The six hyperfine lines of manganese in forsterite has been investigated by the Participant⁵ using nuclear magnetic resonance. By examining the sweep rates between the five neighboring pairs of lines, the scan rates were measured to be the same within 0.7% with the variation random in nature. From time to time this same marker has been used as a field marker and intensity comparison reference for silicon in a manner similar to that employed earlier by the Participant in work⁶ on the rare earths in the fluorites. For the new spectrometer a series of four of these markers with varying intensities was prepared. The design is such that one marker can be inserted uniquely in the back of the cavity and remain unmoved during the running of a series of silicon samples. Several good experimental references^{7,8,9,10} were consulted in connection with these experimental efforts.

Silicon samples were run at X-band at room temperature and at nitrogen temperature (77 K) in an inserted quartz dewar. Helium temperature (4.2 K) dewars are available as is a Heli-tran liquid transfer refrigerator for variable temperature control from helium to room temperature. This system has not been employed at this point in time.

During the ten weeks reported here the system has been put into operation at 77 K and 300 K, has been calibrated, and a number of silicon samples have been run as described in the following. A variety of topics were investigated with an eye toward application of EPR technique. Measurements were attempted in connection with some of them. The topics will be discussed in the following sections.

DANGLING BONDS

A simple model of a (111) surface of a diamond-structure covalent semiconductor suggests one cut bond per surface atom. This leads to the expectation that each such electron would produce a "dangling bond". These electrons, if they remain largely unpaired, should be detectable by EPR measurements. EPR signals have been found from silicon surface regions.¹¹ There has been interest in determining whether this model is correct or whether the atomic rearrangements known to occur on most semiconductor surfaces cause major modifications. The dangling bond concept has been the simple starting point for many theoretical discussions and could not be seriously tested while surface structures remained uncertain. Recently, theoretical computations of surface states have been carried out by Schluter and co-workers¹² for the surface model of Haneman.¹³ This work concluded that paired electrons exist on alternate sites, which would yield only a very weak EPR signal. Recent EPR measurements^{14,15} have shown that there is indeed a negligible EPR signal from well-cleaved silicon surfaces.

In a February 1978 article,¹¹ it was reported that the above-mentioned signal is observed in less well-cleaved surfaces and that the origin is now believed to be due to localized states on microcrack surfaces.

Much data has shown an EPR signal to be present in the surface region of silicon. The results to 1974 have been reviewed.¹⁶ In summary, an EPR signal at $g=2.0055$ of width 6 to 7 gauss, has been found at room temperature from silicon surfaces that have been abraded, produced by crushing¹⁷, by cleavage, irradiated with neutrons, heated in vacuum and quenched¹⁸, and from amorphous films.^{19,20}

Recent work¹⁵ showed that a well-cleaved surface has less than one spin per one hundred atoms. The fact that the spin density from high quality cleaved surfaces can be very much lower than for poorer surfaces points to an origin associated with cleavage rather than with the flat surface. The evidence has led to a conclusion¹¹ that, since the EPR signal that appears when silicon is crushed, cleaved, or abraded is proportional to the areas of microcracks induced, the origin of the signal may be localized states on the surfaces of the microcracks.

In the present work a number of attempts have been made to see so-called dangling bonds. Several likely samples were examined at 77 K and 300 K. The usual surface bonds are tied up with silicon dioxide. Etching treatments have been used in an attempt to free surface bonds. The procedure²¹ used included thirty seconds in HF (49%) followed by thirty seconds in HNO_3 (65%): HF (49%): CH_3COOH = 3:2:2 and rinsing with deionized water. The prescribed treatment²¹ called for 15 minutes in the combined solution but an experience with a 28 milligram sample which vanished in four minutes, reduced the time to thirty seconds in this version.

No detectable resonance was observed in any of these attempts. In one attempt the sample was immersed in liquid nitrogen immediately after etching with the same result.

RADIATION DAMAGE

EPR has been a useful experimental technique for the study of the structure and kinetics of defects in solids. Although restricted to centers that are paramagnetic, under certain conditions it is possible to convert

a defect from the diamagnetic to the paramagnetic state. Semiconductors are especially versatile in this respect since the initial position of the Fermi level is dependent upon the concentration of n- or p- type dopants. In silicon, irradiation produces deep, localized levels which compensate the shallow donor and acceptor levels and consequently pull the Fermi level towards the middle of the bandgap. As this occurs many of the defects alternate between diamagnetic and paramagnetic states.

EPR studies in silicon damage have been particularly successful, as demonstrated by the early series of papers^{22,23,24,25,26} by Watkins and co-workers. Various reviews^{27,28,29} of the subject have appeared over the years. An example of a few of the many silicon centers observed is presented in the following table:

<u>Center</u> *	<u>Host</u>	<u>Doping</u>	<u>Particle</u>	<u>Energy</u>
Si-E	Float zone	n-type; P	electrons	1.5MeV
Si-A(Si-B1)	Pulled	n-type; P	electrons	1.5MeV
Si-N			fast neutrons	Pile spectrum
Si-J(Si-G5)	Pulled	p-type; B,Al,Ga,In	electrons	1.5MeV
Si-C(Si-G7)	Pulled	n-type; P	electrons	1.5MeV

* Note: At least two sets of nomenclature are used in the literature of irradiation induced defect centers.

At this point in time, no irradiated silicon samples have been investigated in this experimental effort.

It is useful to reflect on the feasibility of attempting to examine radiation damaged silicon with the present spectrometer. A number of authors have reported^{14,15,19,30} damage on x-band (9.1 to 9.5 Gigahertz) spectro-

meters at room temperature and liquid nitrogen temperature. However, others have found it advisable to use other experimental conditions for studies of radiation damaged silicon. For example, a number of experimenters^{31,32,33,34} have gone to K-band (19 to 20 Gigahertz) spectrometers. Also work on damaged silicon is often done by using an ENDOR spectrometer.^{35,36,37}

ION IMPLANTED SILICON

Ion implantation is a useful technique for introducing dopants into the silicon crystal. In the process, ions of a particular impurity element, such as boron or phosphorus, are shot from an accelerator into a crystal such as silicon. The choice of the injecting energy determines the depth to which the silicon is doped. However, two major problems limit the desirability of ion implantation. The bombardment with high-energy ions disrupts the orderly arrangement of the crystal lattice giving much surface damage. In addition the ions do not all settle in substitutional lattice sites to become donors or acceptors. This damage lends itself to study by EPR. The first such report³⁸ concerned O^+ ions implanted in Si:Al and Si:B crystals. In this initial report a layer of less than $15,000 \text{ \AA}$ was reported due to 300 KeV O^+ -ions. The so-called Si-P3 center is the dominant paramagnetic defect produced at room temperature in both samples, and it anneals below 200°C . The Si-P1 center dominates above 200°C to near 350°C , where it anneals. Additional EPR investigations led to depth distributions³⁹, other implanted ions,^{40,41} and other defect centers.^{42,43} Another worker has reported⁴⁴ that 1 MeV O^+ -ions implanted in silicon produces amorphous layers which recrystallize after 1000°C anneals.

A more recent study⁴⁵ of interest in the current investigation concerns an EPR investigation into phosphorus implanted silicon. In the investigation, an x-band spectrometer was used at 77°K. It was reported that P-ion implantation reduced the g-factor (magnetogyric ratio). An increase in ion dose from 6×10^{13} to $2 \times 10^{16} \text{ cm}^{-2}$ reduced monotonically the g-factor from 1.9990 to 1.9980 for samples annealed at 1150°C and from 1.9990 to 1.9970 for those annealed at 1070°C. In each case the EPR signal of the conduction electrons was isotropic and the profile was Lorentzian. The line widths increased from 2 gauss to 12 gauss (1150°C) or 30 gauss (1070°C). Successive removal of thin silicon layers resulted in reverse changes; the g-factor gradually returned to 1.9990 and the width to 2 gauss. These results were fit to an analytical expression.

This brief sketch gives an indication of the possibilities that exist with the EPR technique. To this point in time the Participant has not had access to ion implanted samples for investigation with the EPR technique.

NEUTRON TRANSMUTATION DOPING

Boron impurities of the order of 10^{12} cm^{-3} are present in the purest of available silicon. This shallow acceptor has several negative effects on the operation of gallium or indium doped silicon. The boron ionization energy is less than those of the two dopants and the boron levels thermally ionize at lower temperature. This requires additional detector cooling

which increases weight and cost. This difficulty can be eliminated if the boron levels can be populated by carriers from compensating donor levels. Nearly exact compensation of these boron acceptors by the addition of shallow donors can eliminate the difficulty. However, greatly overcompensating the boron is undesirable since this would start to fill the gallium or indium levels. Overcompensation also reduces carrier mobility, carrier lifetime, and detector gain.

A suitable shallow donor is phosphorus, ^{31}P . It can be doped into silicon by neutron transmutation doping. The process is based on converting ^{30}Si to ^{31}Si by an (n, γ) reaction via thermal neutrons from a reactor. The ^{31}Si decays by negative beta-decay to ^{31}P , the impurity sought. The donors produced are distributed uniformly throughout the lattice. Also the rate of production of dopant can be carefully controlled.

These two major advantages of neutron transmutation doping are deflated somewhat by the radiation damage effects that occur during the radiation.⁴⁶ As mentioned in earlier sections, the EPR technique has potential for monitoring the damage and its repair. Several EPR studies of neutron-irradiated silicon have been reported.^{47,48,49} The EPR spectra of phosphorous dopants can be compared to other work^{50,51} on Si:P, giving opportunities for interpretation.

Several pieces of neutron transmutation doped silicon have been examined in this effort. At 300K and 77K EPR spectra were observed. An example of the absorption derivative trace is displayed in Figure 1, on the next page. In the trace one can see a large line and much structure.

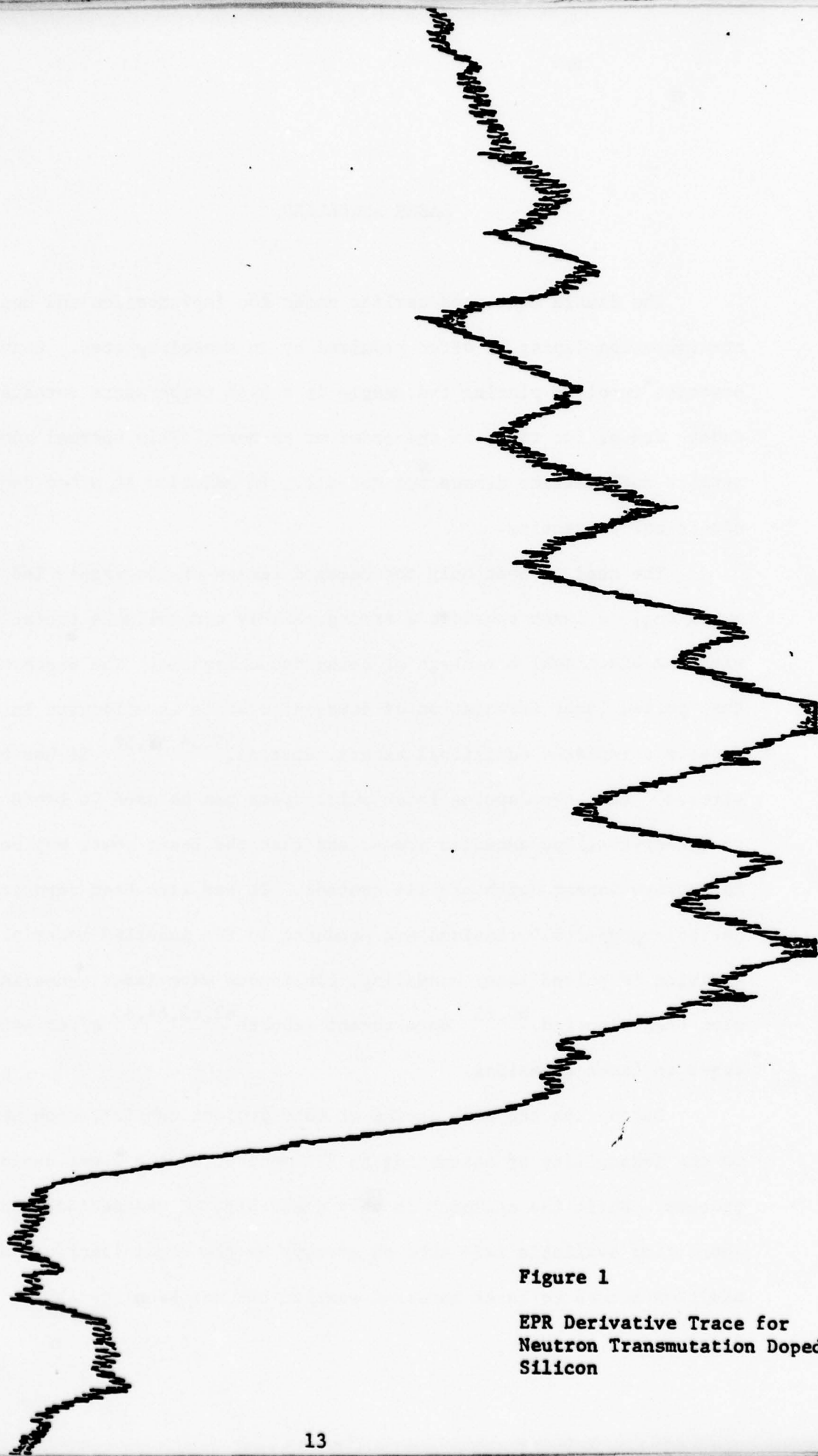


Figure 1

EPR Derivative Trace for
Neutron Transmutation Doped
Silicon

LASER ANNEALING

The damage mentioned earlier under ion implantation and neutron transmutation doping is often repaired by an annealing step. Conventional practice involves placing the sample in a high temperature furnace, perhaps under vacuum, for times on the order of an hour. This thermal annealing repairs much lattice damage but not all. In addition it often degrades electrical properties.

The need to heat only the damaged region of the sample led to laser annealing. A laser provides a strong, highly controllable source of energy with the additional advantage of being monochromatic. The discovery^{52,53,54} that pulsed laser irradiation of damaged crystals is effective in removing defects stimulated additional experimentation.^{55,56,57,58} It has been demonstrated⁵⁷ that overlapping laser pulse spots can be used to produce single-crystalline annealed areas, and that the laser power may be varied to achieve dopant depth profile control. It has also been reported⁵⁹ that periodic property variations are produced in the annealed material. In addition to pulsed laser annealing, continuous wave laser annealing has also been reported.^{60,61} More recent reports^{62,63,64,65} offer other advantages to laser annealing.

During the ten week period of this project consideration was given to the feasibility of attempting an EPR monitor of the laser annealing process. While the approach is very desirable, it was decided that the short time available made such an attempt by the Participant unwise. In addition access to laser annealed samples has not been possible. However

laser annealing continues to appear very attractive because of the potential for defect-free regrowth and because of no loss of dopant during the treatment.

Si:In X-LEVEL

The study of indium-doped silicon for use as a detector in the 3 - 5 micrometer range has led to an interesting puzzle, the identity of the "X-level." A new acceptor level located at 0.111 ± 0.002 eV from the valence band has been observed⁶⁶ in indium-doped silicon. It is believed to be in all Si:In samples except when masked by overcompensation. The existence of this level with an ionization energy significantly less than that of In, 0.156 eV, reduces the maximum temperature for background-limited performance. The usual method of eliminating the effect on low temperature behavior by overcompensating the shallow impurity center is undesirable here due to the degradation in photoconductive properties that would result. Identification of and control of this defect is obviously needed. This has resulted in study^{67,68} of the center, but thus far without identification. Comparisons can be made to earlier work^{69,70} and to more recent studies^{71,72,73,74} on the system. The most promising suggestions for explanations^{66,75} have been that an indium complex is responsible.

During the ten-week summer effort several attempts were made to observe an EPR spectrum in indium-doped silicon. The main results are summarized in the table on the following page.

SUMMARY OF INDIUM-DOPED SILICON RESULTS

Crystal ID	In conc. (cm ⁻³)	X conc. (cm ⁻³)	Mass (mg)	Trace	Observation
DC 002 #18	2.72 X 10 ¹⁷	3.32 X 10 ¹²	28.56	GKM 7-25-78-3	Weak peak near g=2
GZ-163-26	2.04 X 10 ¹⁷	0(undetected)	16.90	GKM 7-13-78-2 GKM 7-25-78-6	Weak peak near g=2
AFML-001-005	2.5 X 10 ¹⁶	0(undetected)	21.59	GKM 7-31-78-11 GKM 7-31-78-15	No spectra 77K or 300K
AFML-001-011			69.10	GKM 8-7-78-8 GKM 8-7-78-9	No spectra 77K or 300K
AFML-001-011			15.27	GKM 8-21-78-1	No spectra 300K
GZ-185 #8	NONE (Gallium impurity)		77.92	GKM 7-6-78-6 GKM 7-6-78-8 GKM 7-24-78-6 GKM 7-25-78-4	Strong peak near g=2 at 77K and 300K

The peak near g=2 is observed in some of the Si:In samples but not in others. It is curiously like a stronger peak observed in a Si:Ga sample, listed last in the table. This spectrum is given in Figure 2 on the next page. The source of this peak has not been identified.

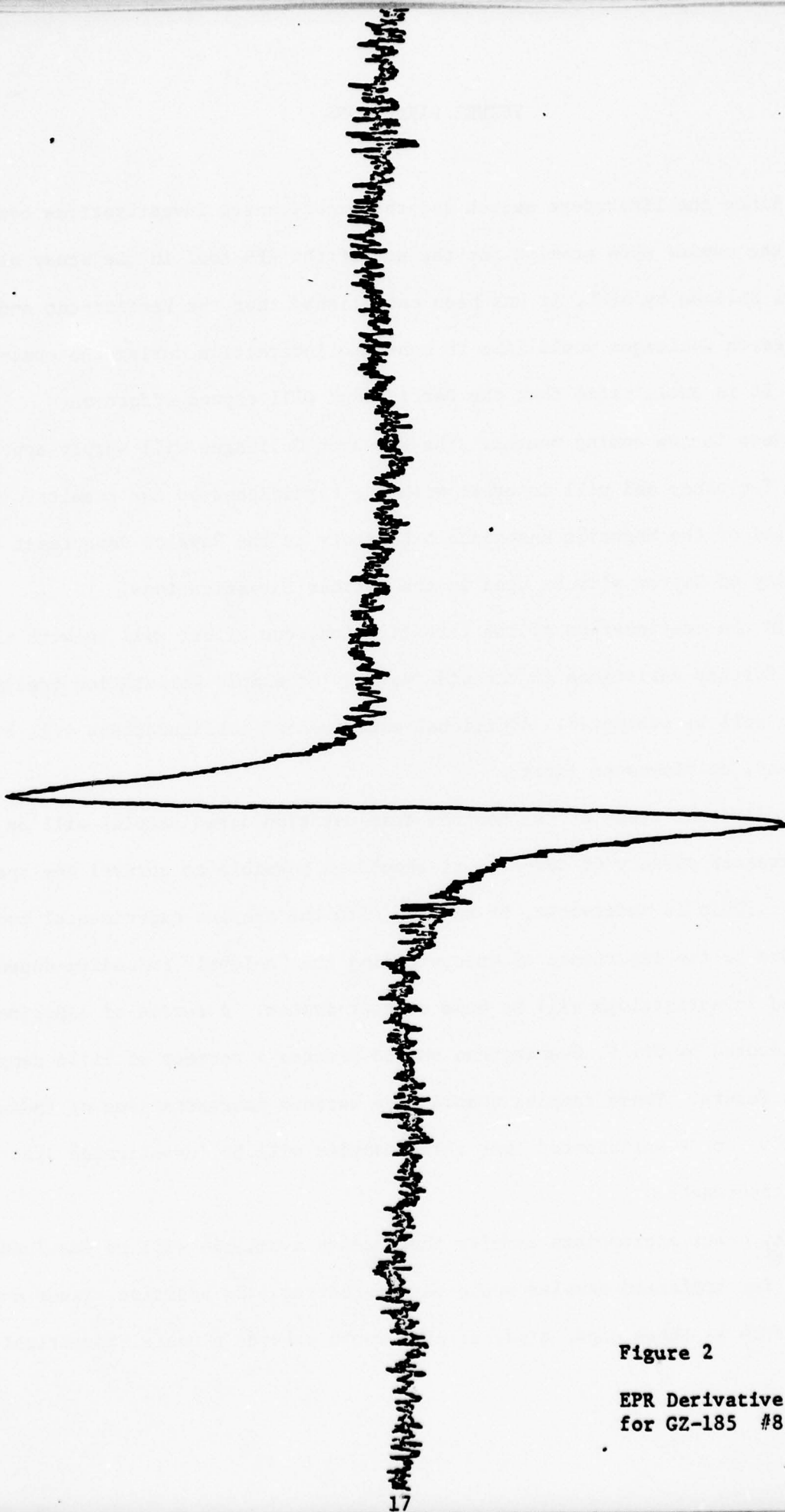


Figure 2

EPR Derivative Trace
for GZ-185 #8

FUTURE DIRECTIONS

Since the literature search and the experimental investigations conducted during the summer show promise for the use of the EPR tool in the study of detector material silicon by AFML, it has been established that the Participant and the Research Colleague would like to continue interaction during the coming year. It is anticipated that the Participant will expend effort on the project in the coming months. The Research Colleague will supply appropriate samples for study and will interact with the Participant on any results. The facilities of the Magnetic Resonance Laboratory in the Physics Department at the University of Dayton will be used in the further investigations.

In the continuation of the investigation, one effort will be with dangling bonds. Further variations in the wide variety of sample preparation treatments possible will be attempted. Additional experimental configurations will be investigated, as discussed later.

Additional study of the neutron transmutation doped samples will be conducted. With a greater variety of samples, it should be possible to unravel the spectrum observed. This is observable, of course, with the present experimental configuration.

Due to the importance of understanding the "x-level" in indium-doped silicon, continued investigations will be made on that system. A series of experiments at AFML conducted by Dr. V. Swaminathan should produce a variety of Si:In samples in the near future. These samples should have various concentrations of indium and "x-level". It is anticipated that these samples will be investigated after a range of heat treatments.

Any other appropriate samples that become available will be examined. For example, ion implanted samples would be interesting. In addition, laser annealed samples such as those under study at AFIT would provide a useful investigation.

The continued EPR measurements will include additional attempts to identify various acceptors and donors in silicon. A guide^{76,77,78} for some acceptors is given in the following table.

Resonance Parameters of Shallor Acceptors in Silicon for a Stress of 700 to 900 kg/cm² Applied Parallel to a Cubic Direction

Acceptor	$g_{ }$	g_{\perp}
B	1.21	2.43
Al	1.18	2.16
Ga	1.14	2.04
In	0.98	1.57

A brief summary of several donors is presented in the following table.

Electron Spin Resonance Spectra of Donors in Silicon

Nucleus	Nuclear spin	g Factor	Hyperfine structure constant G	Line width G
⁷ Li	3/2	1.998	0.3	3.0
³¹ P	1/2	2.000	44	2.9
⁷⁵ As	3/2	2.0004	76	3.6
¹²¹ Sb	5/2	2.000	69	2.7
¹²³ Sb	7/2	2.000	38	2.7

In future investigations, the experimental configuration will be altered. The need is demonstrated in the table taken from Reference 8 and presented on the next page. It is a typical summary of EPR centers seen in semiconductors, mostly silicon. Examination of this table immediately leads to the observation that most of the work is done in the dispersion mode or using the ENDOR technique. The work of this summer was done only in the absorption mode.

Long spin-lattice relaxation times and low defect concentrations place stringent conditions on the experimental techniques required to observe defects in silicon with EPR.⁷⁹ As a result operating conditions and sensitivity are crucial factors in the success of EPR measurements. Because the spin-lattice relaxation times of paramagnetic centers in silicon are very long (as least one second) at low temperatures, it is usually not possible to observe any signal below $\sim 30\text{K}$ in the absorption mode because the signal is completely saturated even at the lowest microwave power levels. Many of the defects in silicon are observed²²⁻²⁶ with EPR under the following conditions:

- 1) Temperature near 20K
- 2) Dispersion mode
- 3) Power of the order of a microwatt
- 4) Modulation frequency less than one kilohertz
- 5) K-band microwaves, and
- 6) A cylindrical cavity.

Table 7.2-2
Examples of Spin Centers and EPR Parameters Studied in Semiconductors

Host crystal	Donor or acceptor		Spin center or impurity				Observed parameters	Mode	Ref.
	Element	Concentration	Treatment	Studied	Concentration per cm ³	Temp, °K			
Silicon	P	10 ¹⁵ -10 ¹⁶	Strained, irradiated (e)	Si-A center	~ 10 ¹⁶	20 and 77	g, hfs, τ	Dispersion	W-7
Silicon	Sb, P, As	10 ¹⁶	Strained	Donors		1.25	Δg , T_2	Dispersion	W-22
Silicon	P	10 ¹⁵ -10 ¹⁶	Strained, irradiated (e)	Si-E center		4.2-155	hfs, g, τ	Dispersion and ENDOR	W-5
Silicon	B, Al, Ga, I		Strained, irradiated (e)	Si-G6 Si-G7		40-110	hfs, g, τ	Dispersion	W-4
Silicon	B, P	~2 x 10 ¹⁶ (B) ~2 x 10 ¹⁶ to 5 x 10 ¹⁶ (P)	Irradiated (e)	Si-G6 formerly Si-A	~ 10 ¹⁶	20.4	hfs, g	Dispersion	C-22
Silicon	B, P	~ 10 ¹⁶	Transition metals diffused in at 1250°C	V ³⁺ Cr ³⁺ Mn ²⁺ Mn ²⁺ Fe ⁰	10 ¹⁶ -10 ¹⁸	1.3 and 20.4	hfs, g	ENDOR	W-28
Silicon			Metals diffused in at 1300°C	Pd and Pt	1-4 x 10 ¹⁶	Up to 20	hfs, g	ENDOR	W-26
Silicon	P, As, Sb	5 x 10 ¹⁶ -6 x 10 ¹⁶	Strained			4.2	hfs, g	Dispersion	J-4 F-12, F-13
Silicon	P	3 x 10 ¹⁶ -10 ¹⁷				12 to 4	hfs, τ		F-2, F-4, F-5
Silicon	¹²¹ Sb, ¹²³ Sb	5 x 10 ¹⁶				1.2	hfs, g	ENDOR	E-6
Silicon	B, As	10 ¹⁵ As, residual B	Neutron irradiated, 10 ¹⁶ -10 ¹⁸ NVT ^a	N(II, III) IX, (I, I') (V, VI), (VII, VIII)		300, 77, 4.2	hfs, g	Absorption	J-9
Silicon	P	6 x 10 ¹⁶ and 8 x 10 ¹⁶				1.3	hfs	ENDOR	J-4
Germanium	P, As	8 x 10 ¹⁴ -5 x 10 ¹⁵				1.3	hfs, ΔH , g		F-8
SiC				S, N, B, Ni	5 x 10 ¹⁷ -10 ¹⁸	78	g, ΔH , hfs		W-19
SiC				B	5 x 10 ¹⁷	14 and 20.4	hfs, g	ENDOR	L-15
SiC				N	3 x 10 ¹⁸				W-27

^aNVT = neutrons per cm³.

Table taken from Reference 8. The references indicated in the table are specified in that text and are not reproduced here.

Under these conditions the spectra are usually observed in fast passage. Usually low temperatures are required because of the need for sensitivity. Unless the sample is intrinsic, the lower temperatures are also required in order to freeze out the carriers so that loading of the cavity is minimized. At these low temperatures even the dispersion signal begins to saturate for higher power levels. For magnetic field modulation frequencies larger than one kilohertz, the lines observed in fast passage behave as though they are partially saturated. K-band frequencies are advantageous over X-band because of the added resolution one gets in the Zeeman spectrum. The cylindrical cavity has a higher Q than the regular rectangular cavity.

In the coming months attempts will be made to alter the spectrometer in this fashion for use in the detection of other paramagnetic centers in silicon. This should make possible new observations in the samples described above.

One of the early reports of use of the ENDOR technique was made by Feher.⁸⁰ He used the approach to investigate shallow donors Sb, P, and As. In addition he studied chemical impurities Bi, Li, and Fe, as well as centers associated with the surface of the sample and with the heat treatment of silicon. Finally he also examined the influence of substitutional germanium atoms on the resonance line in Si:P samples. As time permits this ENDOR technique will be explored.

ACKNOWLEDGEMENTS

The Participant wishes to acknowledge the helpfulness of his Research Colleague, Dr. Patrick Hemenger. Also the interaction with and the aid from Dr. Melvin Ohmer, Dr. V. Swaminathan, Dr. Robert Spry and Steven Smith is appreciated. The Participant acknowledges the interaction with his faculty colleague, Dr. Thomas Graham, and the support of his Department Chairman, Dr. James Schneider.

FOOTNOTES

- ¹"Silicon Monolithic Infrared Detector Array", N. Sclar, R.L. Maddox, and R.A. Florence, Applied Optics 16, 1525 (1977).
- ²"Measurement of High Resistivity Semiconductors Using the van der Pauw Method", Patrick M. Hemenger, Rev. Sci. Instrum. 44, 698 (1973).
- ³"The Growth of Insulating Crystals", J.C. Brice, Rep. Prog. Phys. 40, 567 (1977).
- ⁴"EPR of a Thermally Induced Defect in Silicon", Y.H. Lee, R.L. Kleinhenz, and J.W. Corbett, Applied Physics Letters 31, 142 (1977).
- ⁵"The Preparation and Calibration of a Convenient EPR Field Marker and Intensity Reference", G.K. Miner, T.P. Graham, and G.T. Johnston, Review of Scientific Instruments 43, 1297 (1972).
- ⁶"Effects of a Ce^{3+} Codopant on the Gd^{3+} EPR Spectrum of SrF_2 at Room Temperature", G.K. Miner, T.P. Graham, and G.T. Johnston, Journal of Chemical Physics 57, 1263 (1972).
- ⁷Electron Spin Resonance, Charles P. Poole, Interscience Publishers (1967).
- ⁸Electron Paramagnetic Resonance: Techniques and Applications, Raymond S. Alger, Interscience Publishers (1968).
- ⁹NMR and EPR Spectroscopy, Staff of Varian Associates, Pergamon Press (1960).
- ¹⁰Electron Spin Resonance Spectrometers, T.H. Wilmshurst, Plenum Press (1968).
- ¹¹"Dangling Bonds on Silicon", B.P. Lemke and D. Haneman, Physical Review B 17, 1893 (1978).
- ¹²"Self-consistent Pseudopotential Calculations for Si (111) Surfaces", M. Schluter, J.R. Chelikowsky, S.G. Louie, and M.L. Cohen, Phys. Rev. B 12, 4200 (1975).
- ¹³"Electron Paramagnetic Resonance from Clean Single-Crystal Cleavage Surfaces of Silicon", D. Haneman, Physical Review 170, 705 (1968).
- ¹⁴"New ESR Investigation of the Cleaved-Silicon Surface", D. Kaplan, D. Lépine, Y. Petroff, and P. Thirry, Physical Review Letters 35, 1376 (1975).
- ¹⁵"Low-Temperature EPR Measurements on *in situ* Vacuum-Cleaved Silicon", B.P. Lemke and D. Haneman, Physical Reviews Letters 35, 1379 (1975).
- ¹⁶Japanese Journal of Applied Physics Supplement 2, 371 (1974).

- 17 "Comparison of Thermal Behavior of Vacuum-Crushed, Air-Crushed, and Mechanically Polished Silicon Surfaces by Electron Paramagnetic Resonance", D. Haneman, M.F. Chung, and A. Taloni, Physical Review 170, 719 (1968).
- 18 "Properties of Clean Silicon Surfaces by Paramagnetic Resonance", M.F. Chung and D. Haneman, Journal of Applied Physics 37, 1879 (1966).
- 19 "Structural, Optical, and Electrical Properties of Amorphous Silicon Films", M.H. Brodsky, R.S. Title, K. Weiser, and G.D. Pettit, Physical Review B, 1 2632 (1970).
- 20 "Recombination in Amorphous Semiconductors", R.A. Street, Physical Review B, 17, 3984 (1978).
- 21 "Integrational Etching Methods", Semiconductor Silicon 1977, H.R. Huff and E. Sirtl, eds., The Electrochemical Society, pg. 187.
- 22 "Defects in Irradiated Silicon. I. Electron Spin Resonance of the Si-A Center", G.D. Watkins and J.W. Corbett, Physical Review 121, 1001 (1961).
- 23 "Defects in Irradiated Silicon. II. Infrared Absorption of the Si-A Center", J.W. Corbett, G.D. Watkins, R.M. Chrenko, and R.S. McDonald, Physical Review 121, 1015 (1961).
- 24 "Defects in Irradiated Silicon: Electron Paramagnetic Resonance and Electron Nuclear Double Resonance of the Si-E Center", G.D. Watkins and J.W. Corbett, Physical Review 134, A1359 (1964).
- 25 "Defects in Irradiated Silicon: Electron Paramagnetic Resonance of the Divacancy", G.D. Watkins and J.W. Corbett, Physical Review 138, A543 (1965).
- 26 "Production of Divacancies and Vacancies by Electron Irradiation of Silicon", J.W. Corbett and G.D. Watkins, Physical Review 138, A555 (1965).
- 27 "A Review of ESR Studies in Irradiated Silicon", Radiation Damage in Semiconductors, Academic Press (1964), page 97.
- 28 "EPR Studies of the Lattice Vacancy and Low-Temperature Damage Processes in Silicon", Lattice Defects in Semiconductors 1974, The Institute of Physics, page 1.
- 29 "The Status of Defect Studies in Silicon", Radiation Effects in Semiconductors 1976, N.B. Urli and J.W. Corbett, eds, Institute of Physics, page 1.
- 30 "Electron Paramagnetic Resonance of New Defects in Heavily Phosphorus-Doped Silicon After Electron Irradiation", Radiation Effects in Semiconductors 1976, N.B. Urli and J.W. Corbett, eds, Institute of Physics, page 213.

- 31 "Electron Paramagnetic Resonance of the Neutral ($S=1$) One-Vacancy-Oxygen Center in Irradiated Silicon", K.L. Brower, Physical Review B, 4, 1968 (1971).
- 32 "EPR of a Jahn-Teller Distorted $\langle 111 \rangle$ Carbon Interstitialcy in Irradiated Silicon", K.L. Brower, Physical Review B, 9, 2607 (1974) and Erratum, 4130 (1978).
- 33 "EPR of a $\langle 001 \rangle$ Si Interstitial Complex in Irradiated Silicon", K.L. Brower, Physical Review B, 14, 872 (1976).
- 34 "EPR Observation of the Isolated Interstitial Carbon Atom in Silicon", G.D. Watkins, K.L. Brower, Physical Review Letters, 36, 1329 (1976).
- 35 "Defects in Irradiated Silicon: Electron Paramagnetic Resonance and Electron-Nuclear Double Resonance of the Arsenic- and Antimony-Vacancy Pairs", Edward L. Elkin and G.D. Watkins, Physical Review, 174, 881 (1968).
- 36 "Electron Paramagnetic Resonance of the Aluminum Interstitial in Silicon", K. L. Brower, Physical Review B, 1, 1908 (1970).
- 37 "Defects in Irradiated Silicon: EPR and Electron-nuclear Double Resonance of Interstitial Boron", G. D. Watkins, Physical Review B, 12, 5824 (1975).
- 38 "Electron Paramagnetic Resonance of Defects in Ion-Implanted Silicon", K.L. Brower, F.L. Vook, and J.A. Borders, Applied Physics Letters, 15, 208 (1969).
- 39 "Depth Distribution of EPR Centers in 400-keV O^+ Ion-Implanted Silicon", K.L. Brower, F. L. Vook, and J.A. Borders, Applied Physics Letters, 16, 108 (1970).
- 40 "ESR and Optical Absorption Studies of Ion-Implanted Silicon", B.L. Crowder, R.S. Title, M.H. Brodsky, and G.D. Pettit, Applied Physics Letters, 16, 205 (1970).
- 41 "Electron Paramagnetic Resonance of Ion-Implanted Donors in Silicon", K.L. Brower and J.A. Borders, Applied Physics Letters, 16, 169 (1970).
- 42, 17 O Hyperfine Structure of the Neutral ($S=1$) Vacancy-Oxygen Center in Ion-Implanted Silicon", K.L. Brower, Physical Review B, 5, 4274 (1972).
- 43 "Electron Paramagnetic Resonance of the Lattice Damage on Oxygen-Implanted Silicon", K.L. Brower and Wendland Beezhold, Journal Applied Physics, 43, 3499 (1972).
- 44 "High Energy Ion Implantation", Guenter H. Schwuttke, AFCRL-TR-75-0106 (1975).

- 45 "Electron Spin Resonance of Conduction Electrons in Ion-Implantation-Doped Silicon Layers. Inhomogeneity of the Impurity Distribution", A.Kh. Antonenko, N.N. Gerasimenko, and A.V. Dvurechenskii, Soviet Physics, Semiconductors, 11, 322 (1977).
- 46 "Electrical Studies of Neutron-Irradiated n-Type Si:Defect Structure and Annealing", Herman J. Stein, Physical Review, 163, 801 (1967).
- 47 "Spin-1 Centers in Neutron-Irradiated Silicon", Wun Jung and G.S. Newell, Physical Review, 132, 648 (1963).
- 48 "EPR Studies in Neutron-Irradiated Silicon: A Negative Charge State of a Nonplanar Five-Vacancy Cluster (V_5^-)", Young-Hoon Lee and James W. Corbett, Physical Review B, 8, 2810 (1973).
- 49 "EPR Study of Defects in Neutron-Irradiated Silicon: Quenched-in Alignment Under $\langle 110 \rangle$ -Uniaxial Stress", Young-Hoon Lee and James W. Corbett, Physical Review B, 9, 4351 (1974).
- 50 "Electron Spin Resonance Study of Interacting Donor Clusters in Phosphorus-Doped Silicon at Low Temperatures. I. Shift of Electron Spin Resonance Line at 0.56-4.2K, Yukio Toyoda, Naoki Kishimoto, Koichi Murakami and Kazuo Morigaki, Journal of the Physical Society of Japan, 43, 114 (1977).
- 51 "Electron Spin Resonance Study of Interacting Donor Clusters in Phosphorus-Doped Silicon at Low Temperatures. II. Overhauser Effect and Electron-Nuclear Double Resonance", Yukio Toyoda and Kazuo Morigaki, Journal of the Physical Society of Japan, 43, 118 (1977).
- 52 "Utilization Coefficient of Implanted Impurities in Silicon Layers Subjected to Subsequent Laser Annealing", I.B. Khaibullin, E. I. Shtyrkov, M.M. Zaripov, M.F. Galyautdinov, and G.G. Zakiroy, Soviet Physics, Semiconductors, 11, 190 (1977).
- 53 "Distribution of an Implanted Impurity in Silicon After Laser Annealing", A.Kh. Antonenko, N.N. Gerasimenko, A.V. Dvurechenskii, L. S. Smirnov, and G.M. Tseitlin, Soviet Physics, Semiconductors, 10, 81 (1976).
- 54 "Annealing of Implanted Layers by a Scanning Laser Beam", G.A. Kachurin, E.V. Nidaev, A.V. Khodyachikh, and L.A. Kovaleva, Soviet Physics, Semiconductors, 10, 1128 (1976).
- 55 "Laser Annealing of Boron-Implanted Silicon", R.T. Young, C.W. White, G.J. Clark, J. Narayan, W.H. Christie, M. Murakami, P.W. King, and S.D. Kramer, Applied Physics Letters 32, 139 (1978).
- 56 "Annealing of Te-Implanted GaAs by Ruby Laser Irradiation", J.A. Golovchenko, T.N.C. Venkatesan, Applied Physics Letters, 32, 147 (1978).

- 57 "Spatially Controlled Crystal Regrowth of Ion-Implanted Silicon by Laser Irradiation", G.K. Celler, J.M. Poate and L.C. Kimerling, Applied Physics Letters, 32, 464 (1978).
- 58 "Laser Annealing of Arsenic Implanted Silicon", J. Krynicky, J. Suski and S. Ugniewski, R. Grotzschel, R. Klabes, U. Kreissig and J. Rudiger, Physics Letters, 61A, 181 (1977).
- 59 "Periodic Regrowth Phenomena Produced by Laser Annealing of Ion-Implanted Silicon", H.J. Leamy, G.A. Rozgonyi, and T. T. Sheng, G. K. Celler, Applied Physics Letters, 32, 535 (1978)
- 60 "Laser (cw CO₂)-Induced Electrical Damage in Ge and Si", S.K. Gulati and W.W. Grannemann, Journal of Applied Physics, 48, 3024 (1977).
- 61 "A Laser-Scanning Apparatus for Annealing of Ion-Implantation Damage in Semiconductors", A. Gat and J.F. Gibbons, Applied Physics Letters, 32, 142 (1978).
- 62 "Laser Annealing of Diffusion-Induced Imperfections in Silicon", R.T. Young and J. Narayan, Applied Physics Letters, 33, 14 (1978).
- 63 "Properties of Laser-Assisted Doping in Silicon", K. Affolter, W. Luthy, and M. von Allmen, Applied Physics Letters, 33, 185 (1978)
- 64 "Epitaxial Growth of Deposited Amorphous Layer by Laser Annealing", S.S. Lau, W. F. Tseng, M-A. Nicolet, and J.W. Mayer, R. C. Eckardt and R.J. Wagner, Applied Physics Letters, 33, 130 (1978).
- 65 "Arsenic Diffusion in Silicon Melted by High-Power Nanosecond Laser Pulsing", P. Baeri, S.U. Campisano, G. Foti, and E. Rimini, Applied Physics Letters, 33, 137 (1978).
- 66 "A New Acceptor Level in Indium-Doped Silicon", R. Baron, M.H. Young, J.K. Neeland, and O.J. Marsh, Applied Physics Letters, 30, 594 (1977).
- 67 "Infrared Spectra of New Acceptor Levels in Indium- or Aluminum-Doped Silicon", Walter Scott, Applied Physics Letters, 32, 540 (1978).
- 68 "Interpretation of Hall Measurements", R.D. Larrabee, Presented at the Electrochemical Society Topical Conference on Characterization Techniques for Semiconductor Materials and Devices, Seattle, WA, May 21-26, 1978.
- 69 "Spectroscopic Investigation of Group III Acceptor States in Silicon", A. Onton, P. Fisher, and A.K. Ramdas, Physical Review, 163, 686 (1967).

- 70 "Photoabsorption Cross Section for Silicon Doped with Indium", R.A. Messenger and J.S. Blakemore, Physical Review B, 4, 1873 (1971).
- 71 "Bound-Exciton Absorption in Si:Al, Si:Ga, and Si:In", K.R. Elliott, G.C. Osbourn, D.L. Smith, and T. C. McGill, Physical Review B, 17, 1808 (1978).
- 72 "Edge Luminescence Spectra of Acceptors in Si: Implications for Multiexciton Complexes", S.A. Lyon, D.L. Smith, and T.C. McGill, Physical Review B, 17, 2620 (1978).
- 73 "Annealing Behavior of In Implanted in Si Studied by Perturbed Angular Correlation", E.N. Kaufmann, R. Kalish, R.A. Naumann and S. Lis, Journal of Applied Physics, 48, 3332 (1977).
- 74 "Photoelectric Properties of Indium-Doped Silicon", E.E. Godik and V.P. Sinis, Soviet Physics Semiconductors, 11, 347 (1977).
- 75 Dr. V. Swaminathan, AFML.
- 76 "Paramagnetic Resonance Absorption from Acceptors in Silicon", G. Feher, J.C. Hensel, and E.A. Gere, Physical Review Letters 5, 309 (1960).
- 77 ESR in Semiconductors, G. Lancaster, Plenum Press (1967), page 49.
- 78 "Electron Spin Resonance in Semiconductors", Solid State Physics Volume 13, Seitz and Tumbull, eds., Academic Press (1962), p223.
- 79 "EPR Techniques for Studying Defects in Silicon", K.L. Brower, Review of Scientific Instruments, 48, 135 (1977).
- 80 "Electron Spin Resonance Experiments on Donors in Silicon. I. Electronic Structure of Donors by the Electron Nuclear Double Resonance Technique", G. Feher, Physical Review, 114, 1219 (1959).

A SIMULATION METAMODEL

ROBERT E. YOUNG
Dept. of Industrial Engineering and Operations Research
Wayne State University
Detroit, Michigan 48237

USAF/ASEE Summer Fellow
USAF ICAM Project AFML/LTC
Wright-Patterson Air Force Base, Ohio

The work has established a metamodel of the simulation modeling process. The approach utilizes structured analysis to form a hierarchical cell model which describes the functional decomposition of the modeling process and data flows.

INTRODUCTION

This report contains a metamodel of the simulation modeling process. The metamodel is described from the viewpoint of a simulation modeler and as such reflects the model design development and implementation process. A structured analysis approach (See reference (ROS77)) is used to represent a hierarchical model of the process of developing a model design, implementing the model design as a simulation program, and performing subsequent analysis. The metamodel is taken to two levels completely with several blocks out into lower levels.

A commentary is presented discussing various aspects of the model and a tree structure discussed to help understand how the metamodel's hierarchy is broken out into its components. The metamodel is then presented.

MODEL COMMENTARY

A hierarchy implies levels with priorities associated with each level. Thus, those activities specified at high levels should be considered to have a higher priority or level of importance than those activities at a lower level. This assumption is used to specify activity blocks within a given level and is based upon prevalent attitudes from the literature (see bibliography) and the author's own experience.

The highest level is the A-0 diagram which identifies the model, principle inputs, controls, mechanisms and output. Thus, we have specified the simulation modeling process as separate from its environment except for the inputs, controls, mechanisms, and outputs identified. This assumption is not completely accurate but it is necessary in order to allow the creation of a manageable model.

The entities in A-0 are interfaces between the system (i.e. the simulation modeling process) and its environment. These then are primary external entities of the system and act as primary drivers of the system. These entities are selected as the primary entities in which the inputs and outputs are casually related through a mechanism constrained by controls.

The first level is the A0 diagram which identifies four principal activities, problem formulation, preliminary model formulating, analysis tool development and the actual analysis. The second level is a breakout of each block in A0. Finally, blocks A33 and A34 in diagrams A3 and A4 are broken out to show simulation program creation and simulation model validation. The model hierarchy is presented in a tree diagram (see Figure 1).

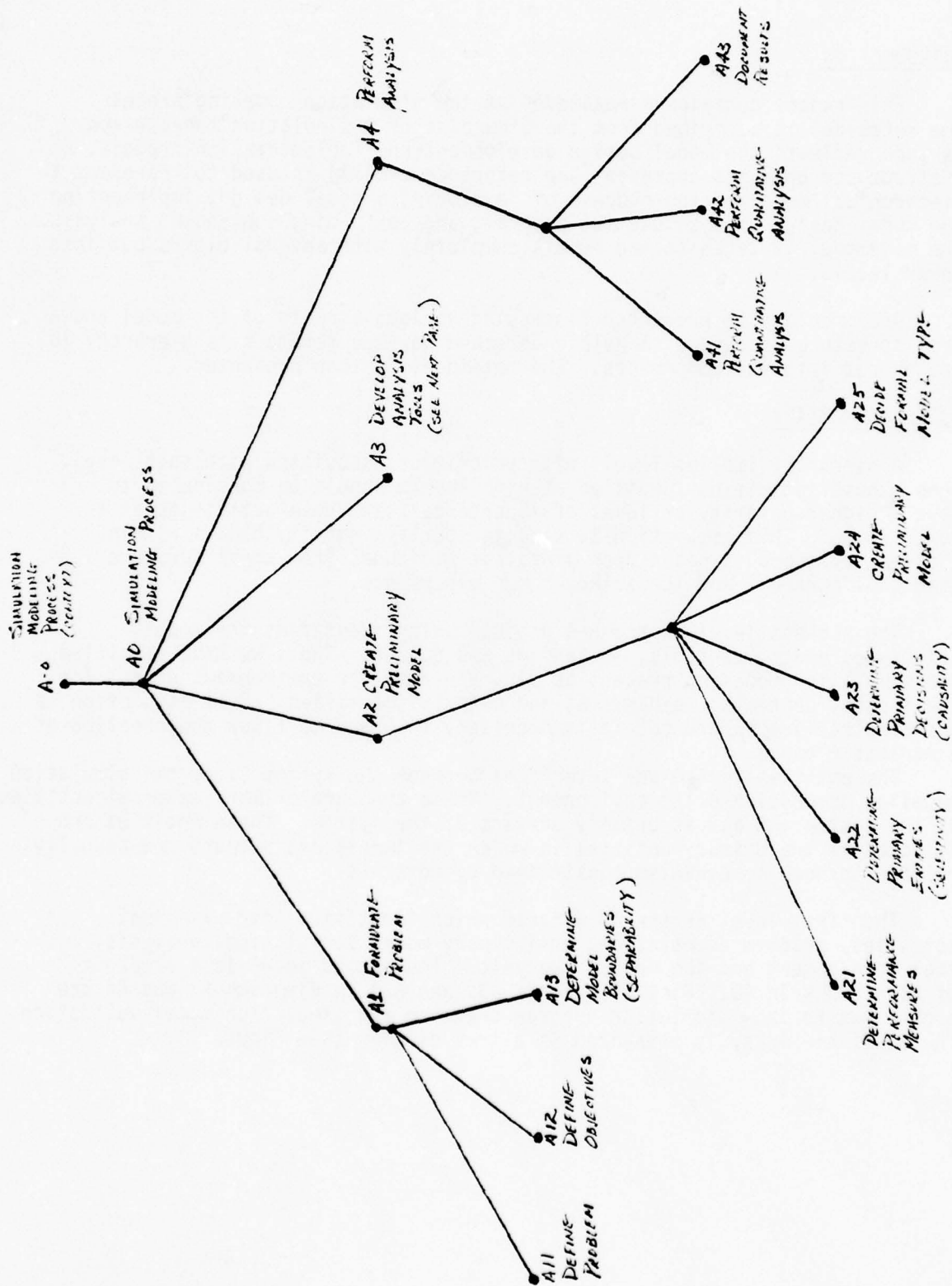


FIGURE 1. ALGORITHM 2. TABLE 2. SIMULATION

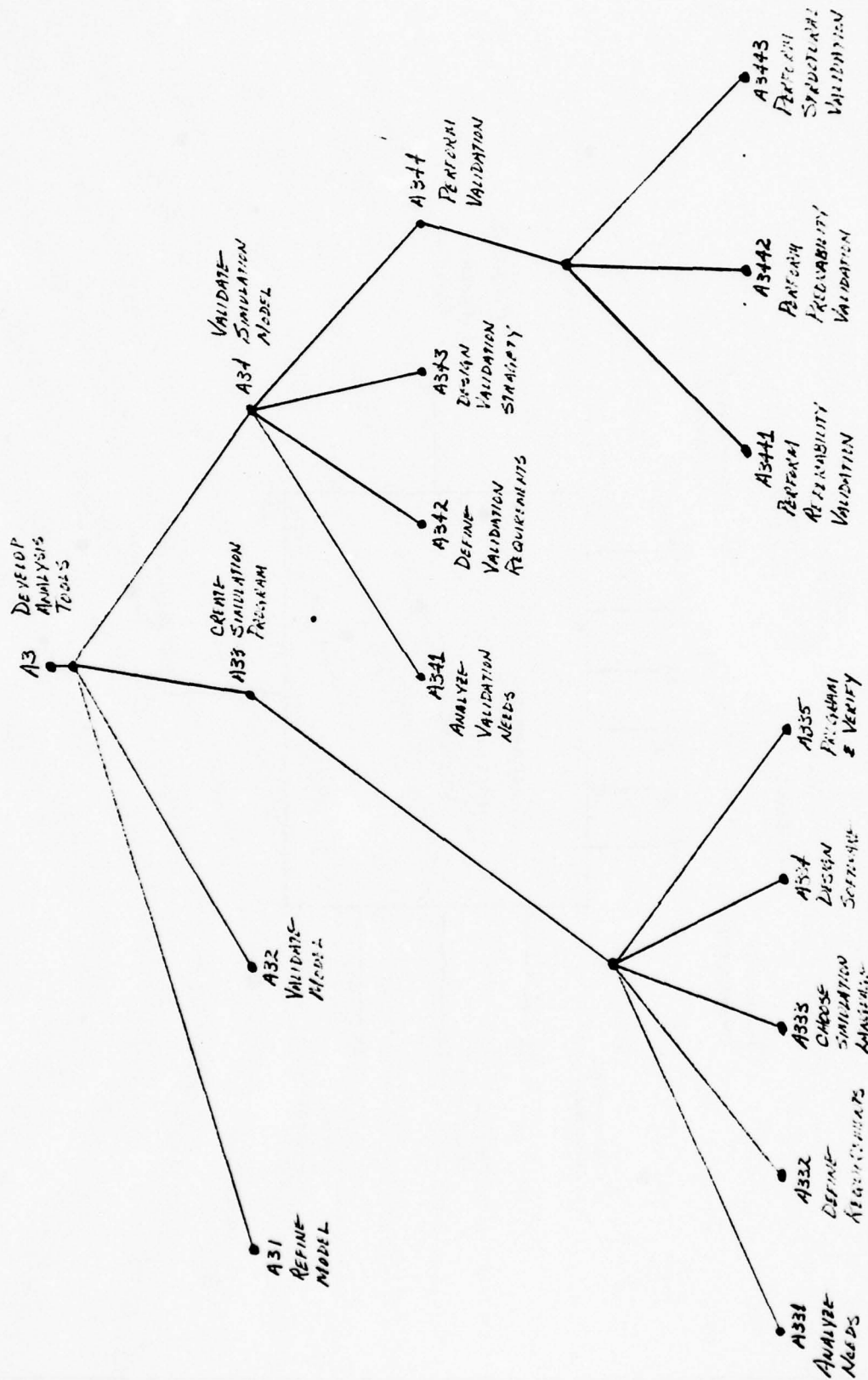
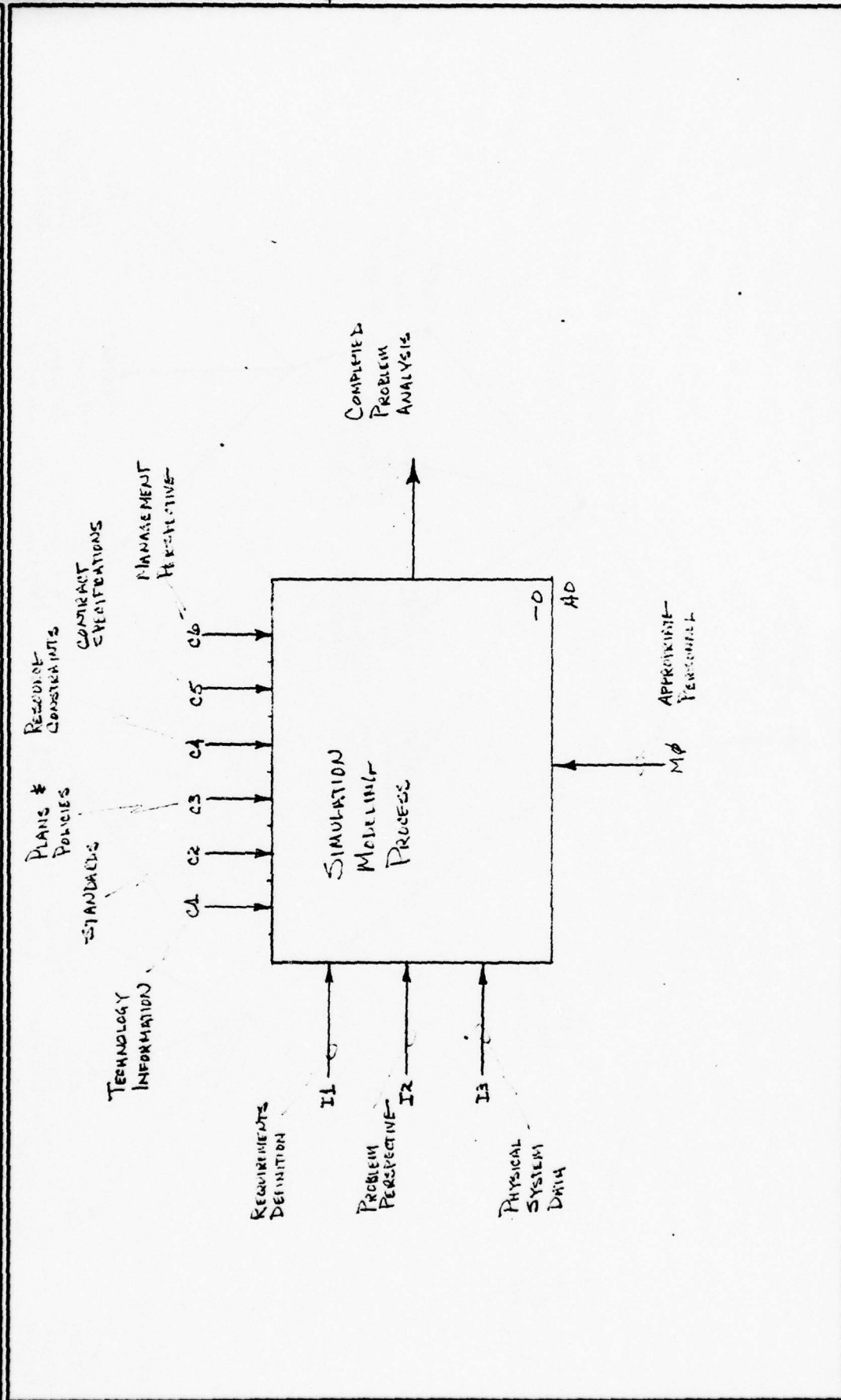


Figure 1 (Cont.) Modeling This Process

USED AT:	AUTHOR: Young	DATE: 8-16-78	WORKING	READER	CONTEXT:
	PROJECT: SIMULATION META-MODEL	REV:	DRAFT		
			RECOMMENDED		
			REPUBLICATION		
NOTES: 1 2 3 4 5 6 7 8 9 10					



NODE: A.0	TITLE: SIMULATION MODELING PROCESS	NUMBER: A.0
-----------	------------------------------------	-------------

GLOSSARY FOR DIAGRAM A-0

Appropriate Personnel

At any point in the modeling process there are two types of personnel associated with a project. The modelers are personnel with experience in modeling and provide continuity as primary drivers of the process. At each functional level and step various personnel interact with the project and modelers to provide insight, critique, etc. These are secondary drivers of the process and are necessary for the primary drivers to perform an adequate job.

Complete Problem Analysis

The final product which includes a verified, validated simulation model, complete statistical analysis of results, and conclusions based upon the model and results, all presented in a well documented report.

Contract Specifications

This refers to contractual agreements which permit the modeling to begin and provide an upper bound based upon costs, and may include personnel, facilities, time, etc.

Management Perspective

The point of view expressed by management or the "client" about the problem, objectives, boundaries, etc. Essentially a management problem perspective. It may or may not be relevant but must be considered.

Physical System Data

Data from the system being modeled. This may be in many forms including numerical, diagramatic, verbal, reports, etc.

Policies and Plans

This refers to plans, organizational policies, and regulations that may govern the software operations of affected organizations. These require general compliance and may originate from the government, contractor, subcontractor, etc.

Problem Perspective

The degree of understanding or insight modelers have about the problem.

Requirement Definition

Provides a statement of need which states why the project is needed, what will be the purpose of the project and justifies the approach to be taken in this project.

Resource Constraints

Refers to constraints placed upon an activity due to availability of manpower, money, time, hardware/software facilities, etc. These are working constraints which operate within the upperbounds set by contract specifications.

Simulation Modeling Process

The act of creating an abstract representation of an actual or proposed system, implementing the representation as a mathematical-logical model and experimentally manipulating it on a digit computer. (PRIT75)

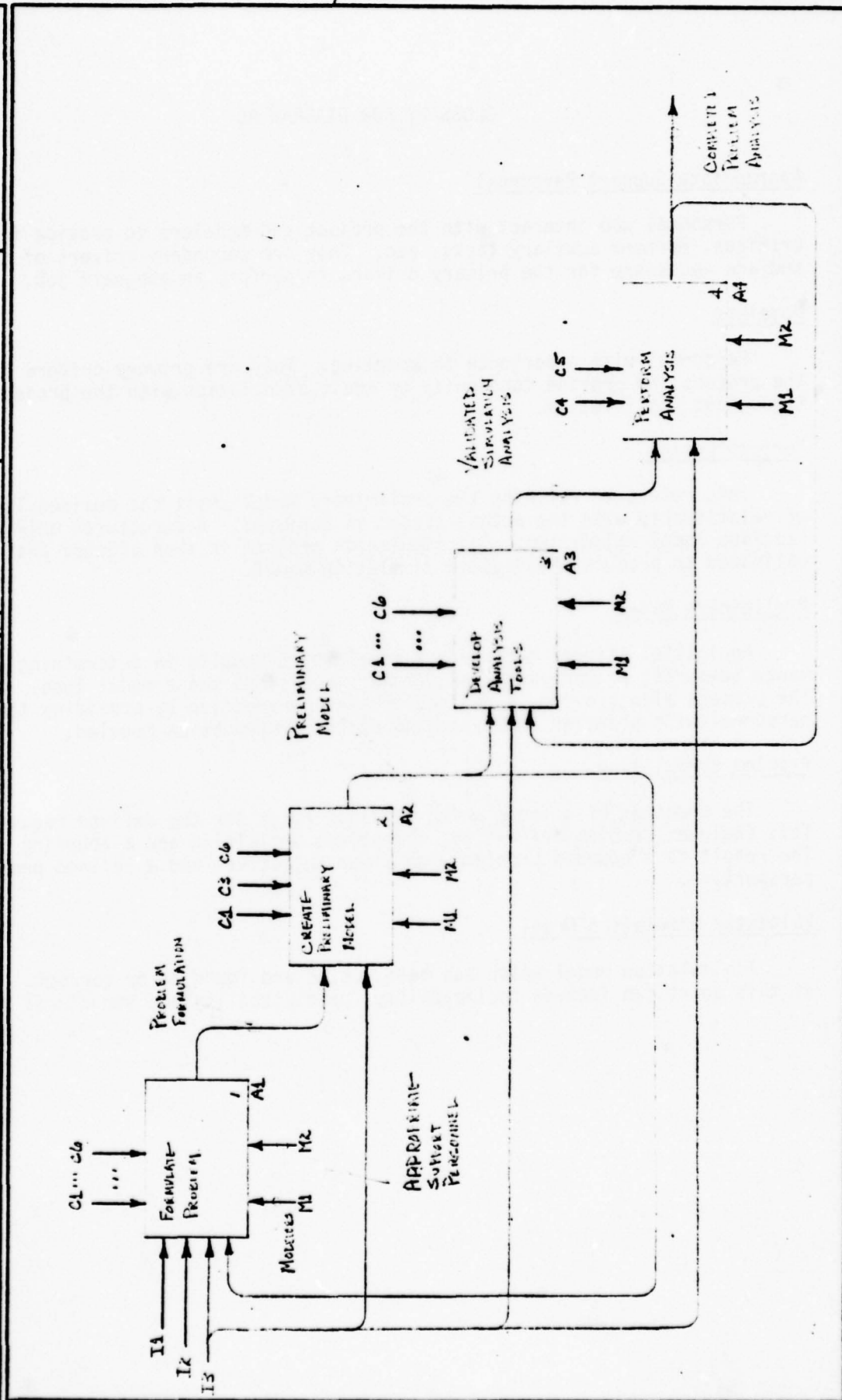
Standards

These refer to existing standards which may require technical compliance.

Technology Information

Refers to state-of-the-art information in pertinent technology and provides guidance on methodology to be used, feasibility of an approach, etc.

USED AT:	AUTHOR: Young	DATE: 8-16-78	WORKING	READER	DATE	CONTEXT:
PROJECT: SIMULATION MODEL	REV:		DRAFT			
NOTES: 1 2 3 4 5 6 7 8 9 10			RECOMMENDED			
			PUBLICATION			



NODE: A0	TITLE: Simulation Modeling Process	NUMBER: A0
----------	------------------------------------	------------

GLOSSARY FOR DIAGRAM A0

Appropriate Support Personnel

Personnel who interact with the project and modelers to provide insight, critique, perform auxiliary tasks, etc. They are secondary drivers of the process and are necessary for the primary drivers to perform an adequate job.

Modelers

Personnel with experience in modeling. They are primary drivers of the process and provide continuity by their association with the project throughout its lifetime.

Perform Design

This refers to refining the preliminary model until the desired level of relationship with the actual system is achieved. A structured walk-through provides model validation. The simulation program is then created and validated to produce a validated simulation model.

Preliminary Model

An initial attempt to create a model which results in determining performance measures, primary entities, primary decisions and a model type. The process also provides a refined problem perspective by providing the personnel with a better understanding of the system being modeled.

Problem Formulation

The creation of a scope which provides focus for the defined requirements. This includes problem definition, objectives definition and a bounding process. The result is a bounded problem with clear objectives and a refined problem perspective.

Validated Simulation Model

A simulation model which has been tested and found to be correct. Validation at this point can include replicability, predictability and structural validity.

Simulation Program

The software embodiment of the validated model which will be exercised on a digital computer.

Validated Model

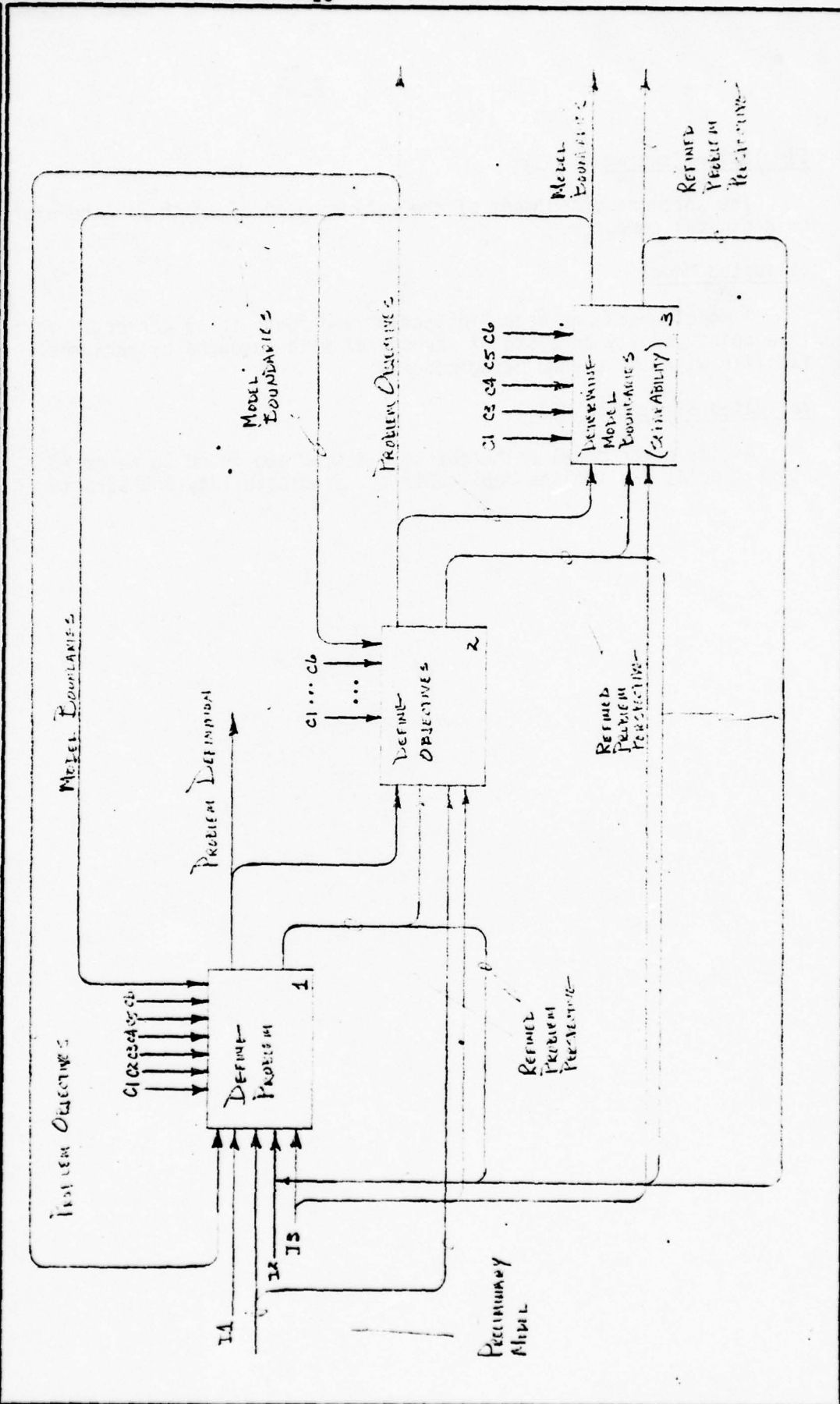
A model which has been "inspected" and found to be correct. Validation at this point usually consists of structural walk-throughs by personnel familiar with the system being modeled.

Validated Simulation Model

A simulation model which has been tested and found to be correct. Validation at this point can include replicability, predictability and structural validity.

SADT[®] DIAGRAM FORM ST008 9/75
Form © 1975 SoftTech, Inc., 460 Totten Pond Road, Waltham, Mass. 02154, USA

USED AT:	AUTHOR: R.L. Young	DATE: 8-16-78	WORKING	READER	DATE	CONTEXT:
	PROJECT: SIMULATION ACTIVITIES	REV:	DRAFT			
			RECOMMENDED			
	NOTES: 1 2 3 4 5 6 7 8 9 10		PUBLICATION			



NODE: A1	TITLE: FORMULA - FORM	NUMBER: A1
----------	-----------------------	------------

GLOSSARY FOR DIAGRAM A1

Model Boundaries

The implementation of the separability assumption. We define the system components as essentially separate from their surroundings creating a "boundary" between our system and its surroundings. This assumes that most of the interactions between the system being modeled and its surroundings can be ignored (separability assumption).

Problem Definition

A definition of the problem or problems which the project will address. This is similar to an objective but does not necessarily have a goal. An objective is a goal which is established from a defined problem.

Problem Objectives

A goal or goals toward which the project will be directed based upon the problem definition. Essentially provides direction to the project.

Refined Problem Perspective

Improved problem understanding or insight derived through constant work or contact with the problem.

GLOSSARY FOR DIAGRAM A2

Decide Formal Model Type

Once a preliminary model is created it can be used in conjunction with the refined problem perspective to decide what type of formal model is appropriate to obtain the project objectives.

Other Model Types

Since this metamodel is only concerned with simulation models we consider any other formal model type as an "other model type".

Performance Measures

A metric which allows a system/model's performance to be quantified.

Primary Decisions

Those relevant decisions which impact the primary entities and are primarily responsible for linking input to output. This is the implementation of the causality assumption (KAR77).

Primary Entities

Those entities which are of primary interest. Specifically, the items which drive the system. This is an application of the selectivity assumption (KAR77).

Simulation Model Type

The formal model which will be used to solve for the objectives is simulation. By simulation we are focusing on stochastic simulation embodied by GPSS, GASP, QGERT, SIMSCRIPT, etc.

GLOSSARY FOR DIAGRAM A3

Model Alterations

These are changes necessitated by deficiencies identified by the analysis process and generally reflect insufficient or incorrect statistical data types or output. However, they are major structural modifications identified by the analysis which must be made in order to reach the objectives.

Model Corrections

These are changes necessitated by deficiencies identified by the validation processes. Usually they do not require major structural changes, but reflect minor inconsistencies within the model.

Qualitative Data

Data which is not numerical in nature. This may include descriptive, graphical, visual, etc.

Refine Model

A refinement of the preliminary model to produce a model which has sufficient detail and structure to allow the objectives to be reached. Model Refinement is accomplished within the structure imposed by a choice of simulation as a formal model type.

Simulation Model Run Time

The actual computer time required to make a "good" run on the computer of the simulation model. A good run is defined as a computer run which produced results with sufficient user confidence to be considered as representative of the system behavior.

Simulation Program

The software embodiment of the validated model which will be exercised on a digital computer.

Validated Model

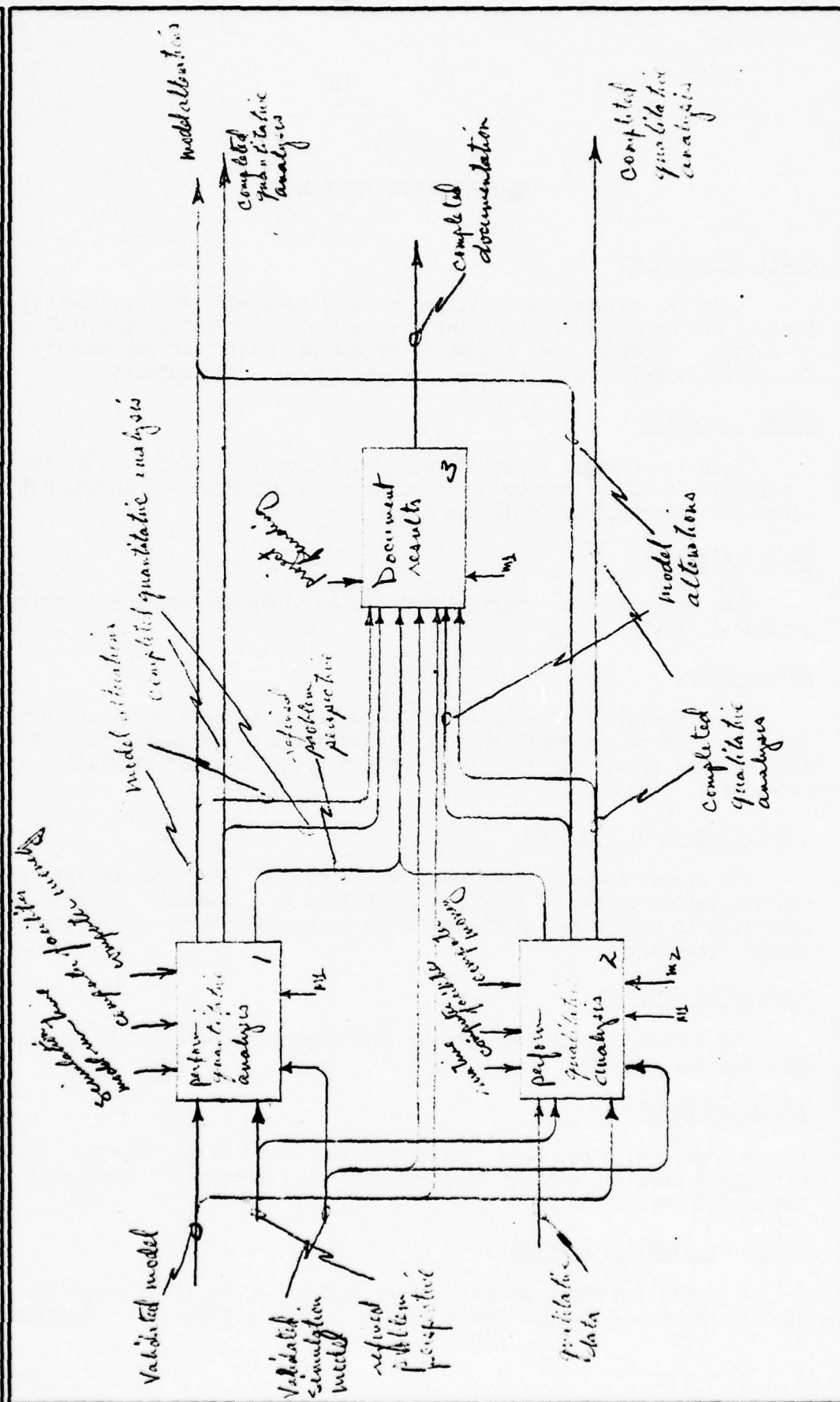
A model which has been "inspected" and found to be correct. Validation at this point usually consists of structural walk throughs by personnel familiar with the system being modeled.

Validated Simulation Model

A simulation model which has been tested and found to be correct. Validation at this point can include replicability, predictability and structural validity.

SADT® DIAGRAM FORM ST098 9/75
 Form © 1975 SofTech, Inc., 460 Totten Pond Road, Waltham, Mass. 02154, USA

USED AT:	AUTHOR: <i>Young</i>	DATE: <i>7-27-76</i>	WORKING	READER	DATE	CONTEXT:
	PROJECT: <i>Simulation model</i>	REV: <i>1</i>	DRAFT			
	NOTES: 1 2 3 4 5 6 7 8 9 10		RECOMMENDED			
			<input checked="" type="checkbox"/> PUBLICATION			



NO:	44	TITLE:	PERFORM ANALYSIS	NUMBER:	44
-----	----	--------	------------------	---------	----

GLOSSARY FOR DIAGRAM A4

Completed Documentation

Fully documented models and analysis. This includes models, programs, approaches, statistical techniques, results, analysis methods, etc.

Computer Facilities

This refers to the physical facilities available to exercise the simulation model and includes both hardware and software. Software attention is directed to statistical analysis. Hardware attention is directed primarily to congestion which impacts throughput time.

Computer Money

The capital available to pay for computer time.

Project Funding

The level of project funding directly impacts the detail and level of documentation.

Qualitative Analysis

The analysis of the physical structure of the system as embodied in the model. This is done in conjunction with the refined problem perspective and identifies flaws in the inherent system structure, understanding of the system synergism and actual or potential problems which may be incidental to the actual project objectives.

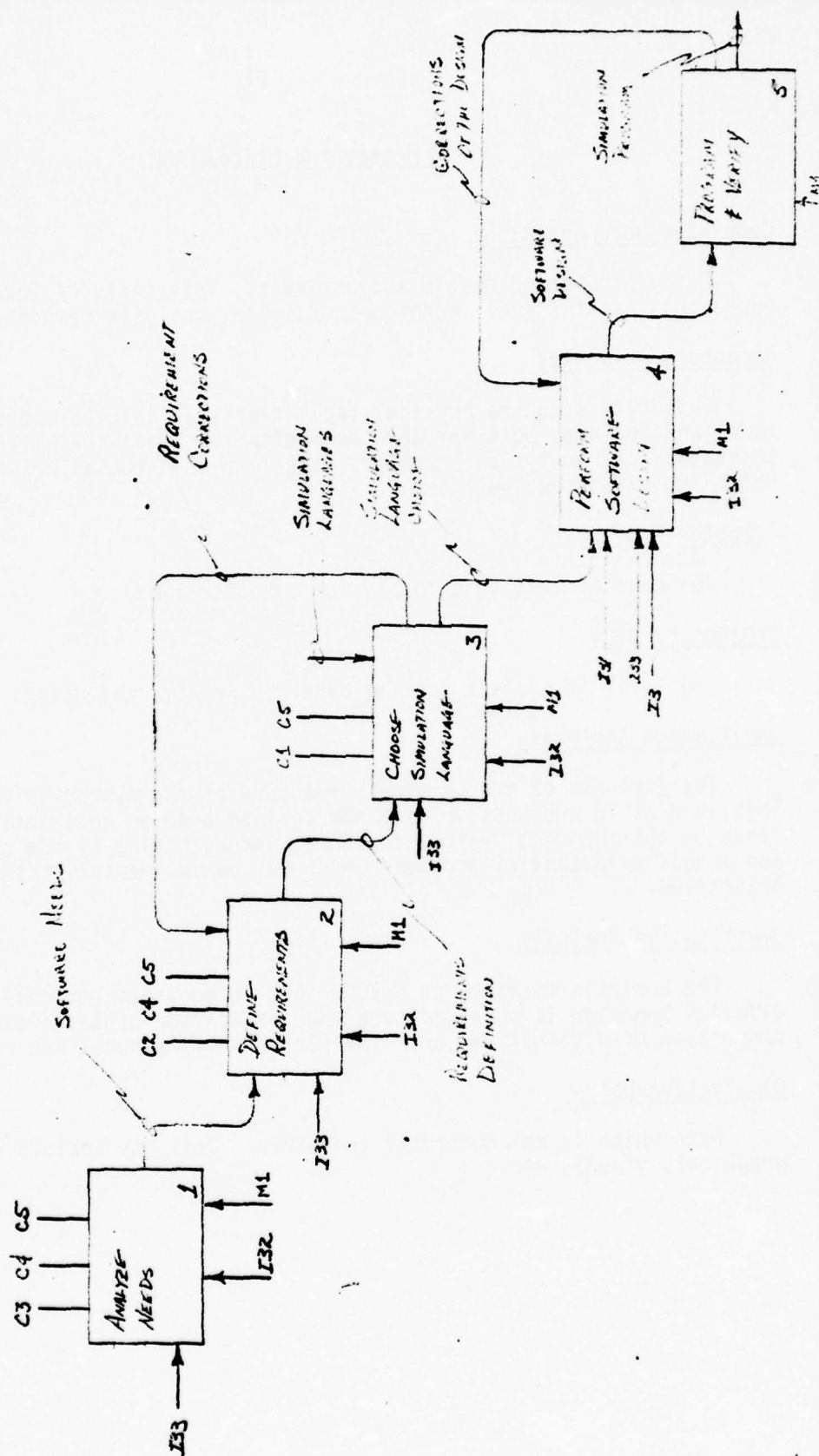
Quantitative Analysis

The analysis required to insure that an accurate estimate of the systems expected behavior is obtained, and the application of techniques which allow comparison of alternatives when the performance measures are random variables.

Qualitative Data

Data which is not numerical in nature. This may include descriptive, graphical, visual, etc.

USED AT:	AUTHOR: Young	DATE: 8-17-78	WORKING	READER	DATE	CONTEXT:
	PROJECT: SIMULATION METAIDEL	REV:	DRAFT			
NOTES: 1 2 3 4 5 6 7 8 9 10			RECOMMENDED			
			PUBLICATION			



NODE: A33	TITLE: CREATE-Simulation Program	NUMBER: A33
-----------	----------------------------------	-------------

GLOSSARY FOR DIAGRAM A33

Analyze Needs

An analysis of the software needs to implement the model.

Choose Simulation Language

A simulation language is chosen whose structure will most readily accept the system model with a minimum of alterations.

Define Requirements

Define the requirements for the model in terms of software implementation based upon defined software needs.

Corrections of the Design

Corrections necessary to produce a verified program.

Program & Verify

Creating the computer program which is the software implementation of the system model and insuring that the computer program is correct.

Requirements Corrections

Corrections for the requirement definitions necessitated because they cannot be met by current simulation languages.

Simulation Languages

Computer languages which provide a structure and terminology to facilitate the building of simulations.

Software Design

This is the creation of detailed design and includes detailed program specifications, data structures, subprogram structure, etc.

GLOSSARY FOR DIAGRAM A34

Perform Validation

The act of validating the simulation program.

Validation Needs

Specific validation needs which have been identified.

Validation Requirements

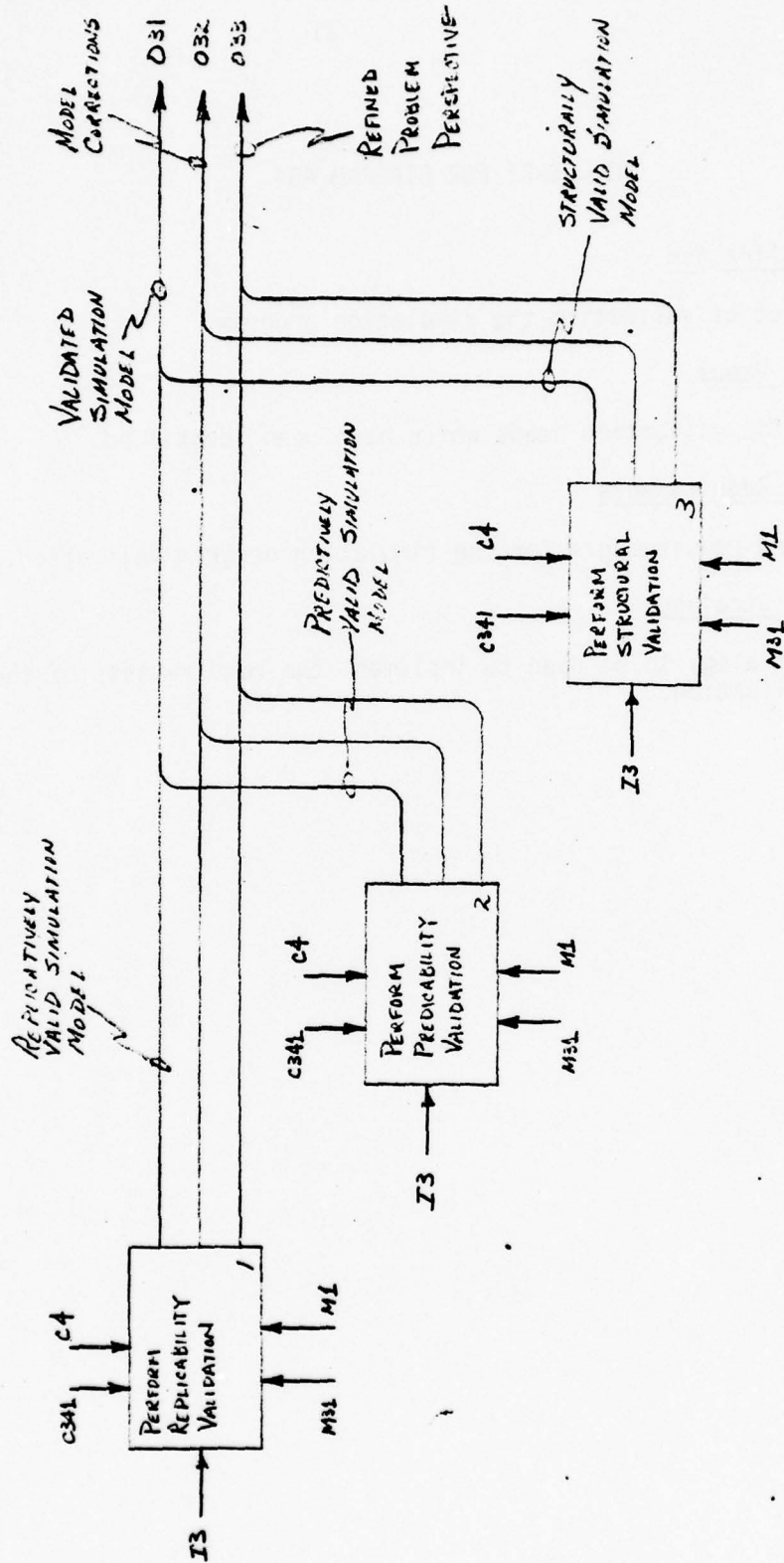
Defined requirements for the simulation program validation.

Validation Strategy

The strategy to be used to implement the requirements of the simulation program validation.

SAOT[®] DIAGRAM FORM ST008 9/75
 Form © 1975 SoTech, Inc., 460 Totten Pond Road, Waltham, Mass. 02154, USA

USED AT:	AUTHOR: <i>R. Young</i>	WORKING	READER	DATE	CONTEXT:
	PROJECT: <i>SIMULATION META MODEL</i>	DRAFT			
		RECOMMENDED			
	NOTES: 1 2 3 4 5 6 7 8 9 10	PUBLICATION			



NODE: *A344*

TITLE: *PERFORM VALIDATION*

NUMBER:

A344

GLOSSARY FOR DIAGRAM A344

Predictively Valid

When data produced from the simulation program accurately predicts data values from the actual system.

Replicatively Valid

When data produced from the simulation program matches historical data acquired from the actual system.

Structurally Valid

Not only does the simulation program reproduce the real system behavior, but it structurally reflects the way in which the real system operates to produce this behavior.

References

- GOR78 Geoffrey Gordon
 System Simulation, 2nd Ed., Prentice-Hall Inc.
 Englewood Cliffs, New Jersey, 1978
- ROS77 Douglas T. Ross
 "Structured Analysis (SA): A Language for
 Communicating Ideas"
 IEEE Transaction on Software Engineering
 Vol. SE-3, No. 1, Jan. 1977, p. 16-35
- MIZ68 J. Mize and J. Cox
 Essentials of Simulation, Prentice-Hall,
 Englewood Cliffs, New Jersey, 1968
- NAY66 Thomas Naylor, Joseph Baliuffy, Donald Burdick
 and Kong Chu
 Computer Simulation Techniques,
 John Wiley & Sons, Inc., New York, 1966
- PRIT75 Alan Pritsker and Robert E. Young
 Simulation with GASP PL/1: A PL/1 Based
 Continuous/Discrete Simulation Language
 John Wiley & Sons, New York, 1975
- SHA75 Robert E. Shannon
 System Simulation: The Art and Science
 Prentice-Hall Inc., Englewood Cliffs, New Jersey,
 1975
- ZEI76 Bernard P. Zeigler
 Theory of Modeling and Simulation
 John Wiley & Sons, New York, 1976

AIR FORCE AVIONICS LABORATORY

Research Associates:

William L. Brogan, University of Nebraska

Thomas P. Graham, University of Dayton

Alfred T. Johnson, Jr., Widener College

Frank L. Pedrotti, Marquette University

ANALYSIS OF THE GPS RECEIVER LOSS-OF-LOCK PROBLEM

William L. Brogan

Participant, USAF-ASEE Summer
Faculty Research Program

Air Force Avionics Laboratory
Wright-Patterson Air Force Base

August 1978

TABLE OF CONTENTS

SECTION	PAGE
1.0 Introduction and Background	1
2.0 Systems Description and Preliminary Considerations Regarding Loss-of-Lock	2
2.1 General System Descriptions	2
2.2 A Preliminary Suggestion	4
2.3 The Present Scheme	4
2.4 Some Possible Approaches to Detecting Loss-of-Lock	5
3.0 A More Detailed Consideration of the Kalman Filter	7
3.1 Definition and Properties of Signals	9
3.2 Statistics of Signals	
4.0 Some Tests for Loss-of-Lock	13
4.1 Tests For Randomness	13
4.2 Simple Tests of Magnitude	14
4.3 Tests For Distribution Parameters	14
4.4 Tests on the Variance of the Δz Signal	16
5.0 Results	17
5.1 A Preliminary Look at Delay Lock Loop Errors	17
5.2 A Preliminary Look at Kalman Filter Residuals	20
6.0 Conclusions and Recommendations	23
References	25
Appendices	
A. Runs Test	26
B. Sequential Likelihood Ratio Test for Detecting Loss-of-Lock	33
C. A Nonsequential Likelihood Ratio Test for the GPS Receiver	40
D. χ^2 Test on the Signal Δz	42

Analysis of the GPS Receiver Loss-of-Lock Problem
William L. Brogan*
Participant, USAF-ASEE Summer
Faculty Research Program

The Global Positioning System (GPS) concept involves the measurement of range to four satellites whose positions are accurately known. This allows the user to determine his own position with high accuracy. The GPS receiver measures these ranges indirectly by means of a measurement of the delay between transmission and receipt of coded satellite signals. Under some conditions the receiver fails to track the satellite signals properly, that is, loses lock. It is not always immediately obvious when this has occurred.

This report presents four general methods which might be of use in detecting loss-of-lock:

- (1) A runs test for testing for randomness of various tracking errors.
- (2) Sequential maximum likelihood ratio test for detection of a valid range signal.
- (3) A non-sequential maximum likelihood test.
- (4) A Chi-square test on signal variances.

All of these methods derive from statistical hypothesis testing, and are intended to apply either to Kalman filter residual signals or to receiver tracking loop errors.

Some of the suggested methods have been illustrated, using hypothetical data. In addition, some preliminary tests have been made on the delay lock loop error and on some Kalman filter residuals from a simulated test of the receiver. Based on these few results, no definite conclusions are drawn. The methods still seem promising, and ought to be given a fuller evaluation when better data becomes available. The methodology that would be useful for further evaluation has been presented in this report.

*Professor, University of Nebraska

1.0 Introduction and Background

The Global Positioning Satellite System (GPS) provides a means for very accurate position determination by a wide class of users. The basic idea is to obtain measurements of the range from the user to several satellites whose positions are accurately known. In principle, simultaneous range measurements to three properly dispersed satellites provide a position fix.

The range measurements are actually mechanized as measurements of signal one-way transit times. This involves the satellite clock and the user clock, and an unknown bias exists between them. The measurements of range are corrupted by this time bias; as a result, they are referred to as pseudo-range measurements. Ideally, four pseudo-range measurements would allow solution for the three unknown position components plus the time bias.

In reality, numerous system errors and signal path distortion factors prevent the ideal single-fix (four measurements) situation from giving satisfactory navigation results. Therefore, a recursive filtering solution (Kalman filter) is used. For a dynamic, maneuvering user such as an aircraft, the result is a so-called integrated navigation system. The name derives from the fact that GPS signals are integrated with another navigation system, usually an inertial navigator.

In an integrated navigation system, the two distinct sources of navigation data are blended together by means of the Kalman filter to obtain the optimal estimate of the system's state at any given time. Simply stated, the advantages of such a system are that the GPS fixes can effectively limit the long-term navigation error growth which is characteristic of inertial systems. Also, the inertial system can be used to velocity-aid the GPS receiver. This means that the apparent shift of the satellite transmitted carrier frequency due to doppler

effects can be greatly reduced or compensated for. This, in turn, permits use of much narrower bandwidth user receivers; hence, the system is less susceptible to jamming*. An earlier study of such a system is found in Ref. [1].

The principle interest in this report is the loss-of-lock problem. The GPS receiver under consideration is the Generalized Development Model (GDM). In simplest terms, it consists of two error-nulling tracking loops. The carrier tracking loop is basically a phase locked loop and the code loop is basically a delay locked loop. The errors in this second loop are errors in the one-way signal transit time and, hence, are proportional to errors in pseudo-range. Because of the large amount of sophisticated signal processing (correlation techniques between the received code and a locally generated code), it is not always obvious when the receiver is not tracking a valid signal. This loss-of-lock phenomenon could occur because of jamming or for other reasons. This report presents several possible approaches to the problem of early recognition of loss-of-lock and other matters related to the recovery from this situation.

2.0 System Description and Preliminary Considerations Regarding Loss-of-Lock.

2.1 General System Description.

A simple functional description of the GPS system is given in Figure 1.

*Reduced sensitivity to jamming is also due, in part, to the highly directional narrow beam antennas and to the spread spectrum techniques used in the receiver.

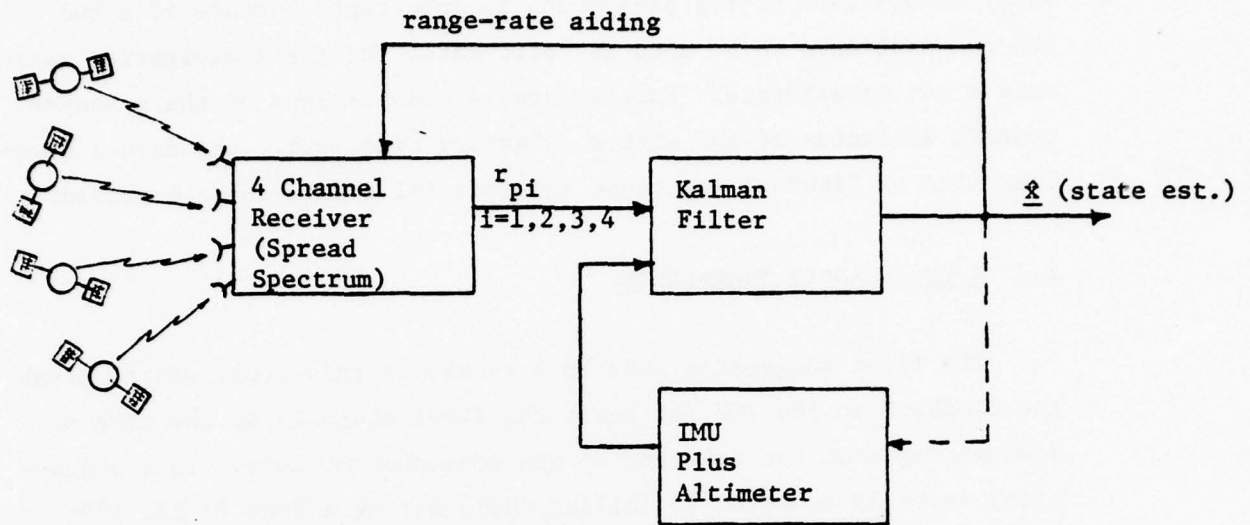


Figure 1: GPS INTEGRATED NAVIGATION SYSTEM

For a detailed description of the receiver operation, Ref. [2 and 3] should be consulted. The radio frequency (RF) signals input to the user antenna are pseudo-random noise code, modulated by high frequency carriers (≈ 1.57 GHz or 1.24 GHz). When the carrier loop is functioning properly, a phase coherent locally generated signal is used to demodulate the received signal. This is the coherent mode. The carrier loop can lose lock because of jamming or excessively high dynamic maneuvers. When this happens, system operation can still continue in a non-coherent mode. In this mode, the delay lock loop continues to operate, giving indications of pseudo-range. The carrier frequency is computed using IMU data (rather than being tracked with the carrier loop).

The problem of interest here concerns the non-coherent mode, and in particular, the loss-of-lock of the delay lock loop, i.e., the code loop. The signals r_{pi} of Figure 1 are measured pseudo-ranges to the i^{th} satellite. When a given channel loses lock, the r_{pi} signal does not normally disappear or show an obvious, abrupt change. Rather, r_{pi} slowly diverges from the correct value, in a random walk fashion. The

early recognition of the phenomenon is important, because if a bad channel continues to be used as valid data, the total navigation estimate \underline{x} can deteriorate. Furthermore, since portions of the \underline{x} vector contain estimates of IMU errors (platform tilt, etc.) the dashed feedback line of Figure 1 can cause the inertial system to be degraded.

2.2 A Preliminary Suggestion.

The first suggestion made as a result of this study was to disable the feedback to the IMU (at least the level channel) at the time of down-moding from the coherent to the non-coherent mode. This suggestion, verbally conveyed to Collins Radio during a June 22-23, 1978, trip to Cedar Rapids, will be analyzed in more detail in later sections. The rationale is that (1) the system is aware when down moding occurs, (2) down moding is probably due to jamming; hence, it raises a flag regarding potential difficulties, (3) by disabling the feedback to the IMU, its integrity is preserved so that the mission can still be carried out. The usual growth of error in inertial systems will commence at this point in time, but is rather slow, especially for an accurately initialized system. The duration of the jamming threat should be short compared with the classical Schuler period, (4) the Kalman filter continues to update all error states and incorporates them into $\hat{\underline{x}}$ for use in receiver aiding and other navigation-dependent mission functions. This would continue to be true until it is definitely decided that loss-of-lock has occurred. At that time, measurements from that channel are no longer processed by the Kalman filter. Even if all channels lose lock, the system could continue in an all-inertial mode.

2.3 The Present Scheme.

At the present time, a reasonableness check is made on each measurement residual before it is processed by the Kalman filter. If the

magnitude of the residual exceeds six sigma, as computed from the filter covariance, that measurement is rejected. This procedure doesn't always give satisfactory performance. The six-sigma threshold is at times too large. (Actually, the six factor is what is too large in my opinion.) In some tests, especially single channel tests, loss-of-lock can occur while the measurement residuals remain sufficiently small to preclude detection by a six-sigma threshold. This could be due to the filter wrongly adjusting certain states to force the residuals to remain small. (This would be less likely in a full four channel configuration.) It could also be due to the very slow rate at which the apparent pseudo-range measurements diverge from truth after loss-of-lock.

The six-sigma threshold is also too small in some cases. If a channel is once rejected, the remaining three channels adjust the state estimates and lead to an overly optimistic estimate of the state uncertainties. The "sigma" part of six sigma becomes so small that when valid measurements from the fourth channel become available, they are consistently rejected.

2.4 Some Possible Approaches to Detecting Loss-of-Lock

There are a number of ways that might be considered as a means of detecting loss-of-lock. They differ in the approaches, assumptions made and signals used. Four approaches mentioned in Ref. [4] are:

(1) A priori signal strength estimator. A measurement of the RF-signal strength would be used, along with estimated antenna gain and the output of the automatic gain control circuitry (AGC). The jamming power to signal power ratio can be estimated. If this ratio exceeds known limits, it is quite certain that loss-of-lock has occurred or soon will. This RF method will not be pursued further in this study. This does not

constitute a value judgement of the method. In fact, it could prove to be a simple and reliable method. Also, it could potentially indicate when jamming has ceased and when reacquisition should be attempted.

(2) Position Variance Monitor. This method is similar to the present scheme, except it could be applied to some other signal (delay lock loop error, for example) and the method of determining the acceptable variance might differ.

(3) Signature Detection. This refers to a generic class of methods which depend on some characteristic behavior of signals in the delay lock loop during loss-of-lock. The methods and characteristics are "as yet unknown".

(4) Signal Presence Monitor. This title refers to classical detection theoretic methods.

Concurrently and independently of Ref. [4], several methods of approaching the loss-of-lock detection problem have been considered in this study. These methods, to be presented, may all fit one or the other of the above categories, but the correct assignment to a given category is not always obvious.

All of the following approaches grow out of statistical hypothesis testing in one form or another. Some of the relevant considerations are: What signals should be used? Are they easily accessible? What are their data rates? What levels of confidence are desired on the decisions? What are the implementation costs of a given scheme? During jamming, receiver bandwidths are narrowed, giving long response times. As a result, measurement noise at five-second intervals is clearly not uncorrelated. This means that there are system states that are being ignored. What is the effect of these unestimated states?

By choice, consideration in the study is limited to receiver output signals (available once per second) or to Kalman filter signals (available once per five seconds.) Some of the tests suggested would apply also to the receiver's delay lock loop tracking error (available every 20 milliseconds), but it seems more difficult to get at signals such as this, buried inside the receiver. The questions of data rates and levels of confidence are intimately tied to the question of how long one can afford to wait before detecting a loss-of-lock. The receiver's output measurements, at a one-per-second rate, should be able to detect loss-of-lock much earlier than the Kalman signals, for a given level of confidence. Also, it intuitively seems wasteful to throw away four receiver measurements out of every five as is now the case.

Another approach to the validation of range measurements, which has been tested on the tracking of deep space shots [5] is to assume that range residuals form a polynomial in time. A Kalman filter is used to estimate the coefficients of the polynomial, and to provide an estimate of the residual variance. This variance is then used in a magnitude test similar to (2) above and Section 4.2 to follow. Although of general interest, this method does not seem to be a reasonable approach in the present application. It seems doubtful that a new Kalman filter is desirable to validate inputs to the original Kalman filter. If the computer space were available for extra states, some of the ideas [5] could be incorporated into the filter, and the problem of neglected measurement noise correlations could be corrected at the same time. This would be done by providing a better model of range residual behavior.

3.0 A More Detailed Consideration of the Kalman Filter

When the recommendation of Section 2.2 is implemented, a more detailed representation of the Kalman filter portion of Figure 1 can be represented as in Figure 2. That is, the dashed feedback signal path is now open so that there is no control input to the IMU.

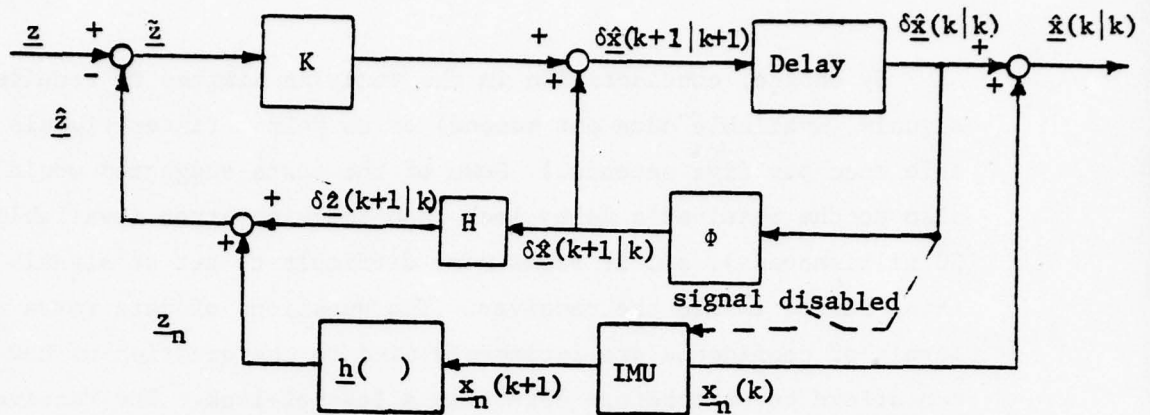


FIGURE 2: BLOCK DIAGRAM OF THE KALMAN FILTER

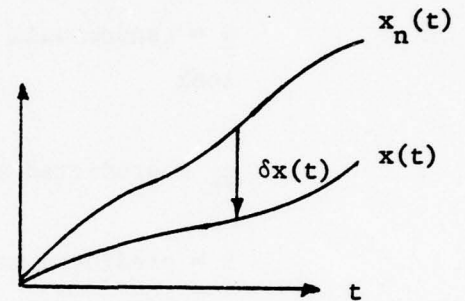
The input \underline{z} represents the set of four pseudo-range measurements from the receiver, corrupted by additive measurement noise \underline{v} . When loss-of-lock occurs in the code loop, the receiver does not suddenly shift its output from $\underline{z} = \underline{r}_p + \underline{v}$ to just noise \underline{v} . Rather, because of the way the code loop discriminator works, the output gradually drifts off from its value before losing lock. This can be modeled as $\underline{z} = \underline{w} + \underline{v}$ where \underline{w} is a noise or error term which typically would diverge from \underline{r}_p after loss of lock. It can probably be modeled as a random walk.

Simulations and lab tests have shown that, at least in the single channel case, when code loop loss-of-lock occurs, $\underline{\hat{z}}$ will still track \underline{z} fairly tightly, i.e., the residual $\underline{\hat{z}}$ is maintained at a small magnitude. This is caused by the "corrections" being applied to the IMU (in the usual case of full feedback control) causing the IMU parameters and/or clock parameters to take on whatever wrong values are necessary to keep the measurement residuals small. This means that pseudo-range residuals may not be good signals for use in detecting loss-of-lock. What is needed is an independent source of pseudo-range estimates which can be compared with receiver outputs. The IMU/clock system can provide these independent estimates, provided that GPS signals are not used to recalibrate the IMU. This decoupled concept, as shown in Figure 2, is the suggestion made in Section 2.2. It applies whenever the receiver goes from the coherent to the non-coherent mode. No significant reduction in

navigation accuracy would be expected because of the opened feedback path. The error states $\delta \underline{\hat{x}}$ continue to be estimated (from GPS signals) and used in the whole value state estimates $\underline{\hat{x}} = \underline{x}_{nav.} + \delta \underline{\hat{x}}$ (feed forward control). This situation would prevail until it is clearly established that loss-of-lock has occurred. Thereafter, the pure inertial system output $\underline{x}_{nav.}$ would be used. The remainder of this report examines the implications of the scheme just proposed.

3.1 Definition and Properties of Signals in (and Related to) Fig. 2.

Referring to Fig. 2, the following definitions and relationships are states without comment. These will be useful in the sequel.



\underline{x} = True State

\underline{x}_n = IMU estimate of \underline{x}

$\delta \underline{x}$ = Error in \underline{x}_n , i.e., $\delta \underline{x} = \underline{x} - \underline{x}_n$ or $\underline{x} = \underline{x}_n + \delta \underline{x}$

$\delta \underline{\hat{x}}$ = Filter estimate of $\delta \underline{x}$; $\delta \underline{\hat{x}} = \underline{\hat{x}} - \underline{x}_n$

$\underline{\hat{x}} = \underline{x}_n + \delta \underline{\hat{x}}$ = est. of total state

$\delta \underline{\tilde{x}} = \text{error in } \delta \underline{\hat{x}} = \delta \underline{x} - \delta \underline{\hat{x}}$

$\underline{\tilde{x}} = \text{error in } \underline{\hat{x}} = \underline{x} - \underline{\hat{x}} = \underline{x}_n + \delta \underline{x} - (\underline{x}_n + \delta \underline{\hat{x}}) = \delta \underline{x} - \delta \underline{\hat{x}}$

$\therefore \delta \underline{\tilde{x}} \equiv \underline{\tilde{x}}$

z = actual measurement. Although there may be a vector set of measurements, they are processed one at a time. Therefore, all measurement quantities may be treated as scalars.

$$z = h(\underline{x}) + \underline{v} = h(\underline{x}_n + \delta \underline{x}) + \underline{v} = h(\underline{x}_n) + H\delta \underline{x} + \underline{v} \quad \text{when locked on}$$

$$z = \underline{w} + \underline{v} \quad \text{when not locked on}$$

\underline{v} = zero mean random meas. noise

\underline{w} = random walk variable; initial value is range at time of loss-of-lock

$$\underline{z}_n = \text{predicted meas. based only on IMU} = h(\underline{x}_n)$$

$$\begin{aligned} \hat{z} &= \text{predicted meas. based on IMU + GPS} = h(\hat{\underline{x}}) \\ &= h(\underline{x}_n + \delta \hat{\underline{x}}) = h(\underline{x}_n) + H\delta \hat{\underline{x}} = \underline{z}_n + H\delta \hat{\underline{x}} \end{aligned}$$

$$\begin{aligned} \tilde{z} &= z - \hat{z} = \text{meas. residual} = h(\underline{x}_n) + H\delta \underline{x} + \underline{v} - h(\underline{x}_n) - H\delta \hat{\underline{x}} \\ &= H(\delta \underline{x} - \delta \hat{\underline{x}}) + \underline{v} = H\tilde{\underline{x}} + \underline{v} \end{aligned}$$

$$\begin{aligned} \Delta z &= z - \underline{z}_n = h(\underline{x}_n) + H\delta \underline{x} + \underline{v} - h(\underline{x}_n) \\ &= H\delta \underline{x} + \underline{v} = \underline{z} + H\delta \underline{x} \end{aligned}$$

} when
locked
on

$$\tilde{z} = \underline{w} + \underline{v} - \hat{z} = \underline{w} + \underline{v} - \underline{z}_n - H\delta \hat{\underline{x}}$$

$$\Delta z = \underline{w} + \underline{v} - \underline{z}_n = \tilde{z} + H\delta \hat{\underline{x}}$$

} when not
locked on

3.2 Statistics of Signals: Making use of the relations in Section 3.1, the means and covariances of the principal signals of Figure 2 are as follows:

$$E(\delta \hat{\underline{x}}) = E\hat{\underline{x}} - E\underline{x}_n = \underline{x} - \underline{x}_n = \delta \underline{x}$$

$$\text{cov}(\delta \hat{\underline{x}}) = P \text{ (Kalman covariance)}$$

$E(\delta \underline{x}) = \delta \underline{x}$ (This means that IMU errors are treated as deterministic)

$$E(\hat{\underline{x}}) = \underline{x}_n + E(\delta \hat{\underline{x}}) = \underline{x}_n + \delta \underline{x} = \underline{x}$$

$$\text{cov}(\hat{\underline{x}}) = E\{(\underline{x} - \hat{\underline{x}})(\underline{x} - \hat{\underline{x}})^T\} = \text{cov}(\underline{x}) = \text{cov}(\delta \hat{\underline{x}}) = \text{cov}(\delta \underline{x}) = P$$

$$E(\tilde{\underline{x}}) = 0$$

$$E(\tilde{\underline{z}}) = H E \tilde{\underline{x}} + E v = 0 \text{ when locked on}$$

$$\left. \begin{aligned} E(\tilde{\underline{z}}) &= E(\underline{w}) - E(\underline{z}_n + H \delta \hat{\underline{x}}) \\ &= E(\underline{w}) - h(\underline{x}) \\ &= h(\underline{x}_0) - h(\underline{x}) \end{aligned} \right\} \text{when not locked on}$$

\underline{x}_0 = state (true) at time lock is lost

$$E(\Delta \underline{z}) = E \tilde{\underline{z}} + H \delta \underline{x} \text{ in both cases, but } \tilde{\underline{z}} \text{ differs.}$$

$$\begin{aligned} \text{cov}(\tilde{\underline{z}}) &= H P H^T + R \text{ when locked on} \\ &= H P H^T + R + \sigma_n^2 (t_k - t_0) \text{ when not locked on} \end{aligned}$$

$\text{cov}(\Delta \underline{z}) = \text{cov}(\tilde{\underline{z}}) + H P H^T + \text{cross covariance terms (closer analysis follows). When locked on}$

$$\Delta \underline{z} - E(\Delta \underline{z}) = H \delta \underline{x} + \underline{v} - H E(\delta \underline{x}) - E(\underline{v}) = \underline{v} - \bar{\underline{v}} \text{ (if } \delta \underline{x} \text{ is treated as deterministic so that } E \delta \underline{x} = \delta \underline{x})$$

$\text{cov}(\Delta \underline{z}) = R$, if $\delta \underline{x}$ is treated as deterministic error. (Actually this term is $R + H P_{\text{nav}} H^T$, where P_{nav} is the cov. of IMU errors $\delta \underline{x}$.) The simpler result is consistent with many other steps which also assumes $\delta \underline{x}$ is a deterministic error growth. The same result follows when the $\Delta \underline{z} = \tilde{\underline{z}} + H \delta \hat{\underline{x}}$ form is used, but one must note and use $E \delta \hat{\underline{x}} = \delta \underline{x}$ and $\delta \hat{\underline{x}} - \delta \underline{x} = -\tilde{\underline{x}}$ and use the cross covariances between $\tilde{\underline{z}}$ and $\tilde{\underline{x}}$.

When not locked on

$$\begin{aligned}\Delta z - E(\Delta z) &= \underline{w} + \underline{v} - \underline{z}_n - (\underline{\bar{w}} + \underline{\bar{v}} - \underline{\bar{z}}_n) \\ &= (\underline{w} - \underline{\bar{w}}) + (\underline{v} - \underline{\bar{v}})\end{aligned}$$

$$\text{So cov } (\Delta z) = \sigma_n^2 (t - t_0) + R$$

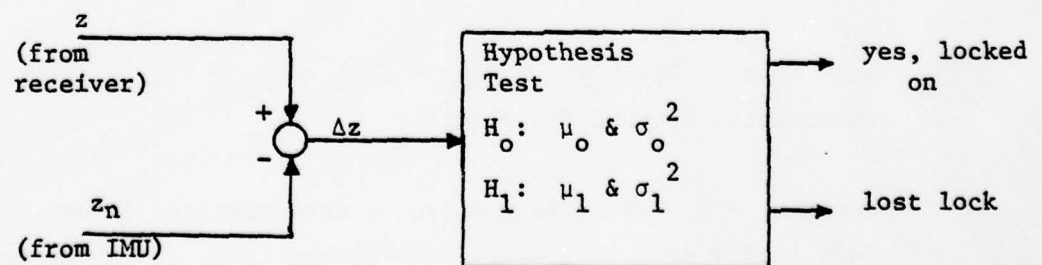
Recap:

$$\begin{aligned}\text{When locked-on: } E(\Delta z) &= H\delta \underline{x} \\ (H_1:) \quad \text{cov } (\Delta z) &= R\end{aligned}$$

$$\begin{aligned}\text{When not locked on: } E(\Delta z) &\approx h(\underline{x}_0) - h(x(t)) + H\delta \underline{x} \\ (H_0:) \quad &\approx \dot{r}(t-t_0) + H\delta \underline{x} \\ \text{cov } (\Delta z) &= R + \sigma_n^2 (t - t_0)\end{aligned}$$

Therefore, the differences between being locked on and not being locked on are linear functions of time for both signals $\mu = E(\Delta z)$ and $\sigma^2 = \text{cov } (\Delta z)$ (not the same linear functions, however).

Because of the linear growing differences between H_0 and H_1 , a hypothesis test on variance and/or mean is suggested as a possible method to determine loss-of-lock.



One such approach is presented in Appendix D.

AD-A065 650

OHIO STATE UNIV RESEARCH FOUNDATION COLUMBUS
USAF-ASEE (1978) SUMMER FACULTY RESEARCH PROGRAM (WPAFB). VOLUM--ETC(U)
NOV 78 C D BAILEY

F/G 1/3

F44620-76-C-0052

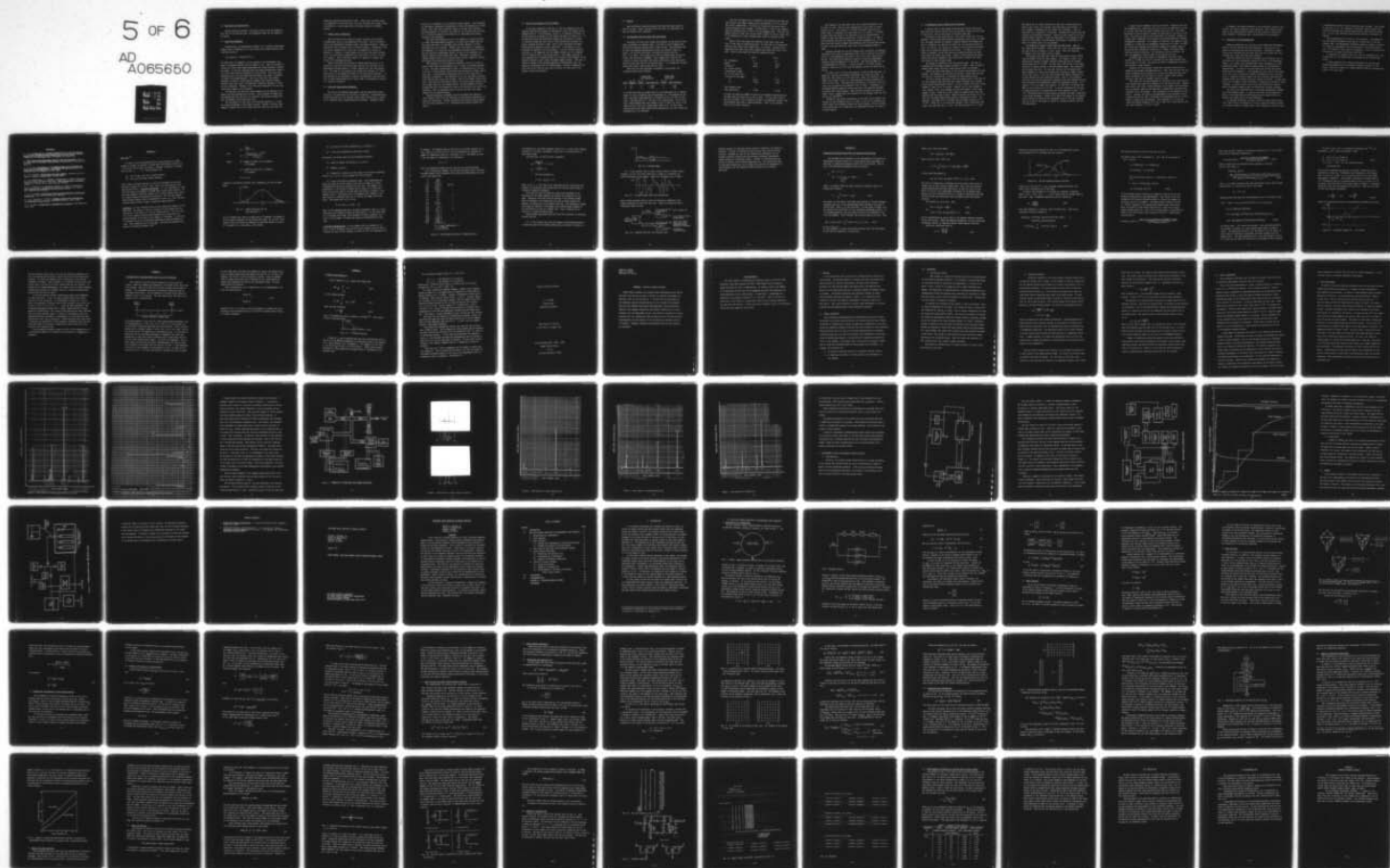
UNCLASSIFIED

AFOSR-TR-79-0231

NL

5 OF 6

AD
A065650



4.0 Some Tests For Loss-of-Lock

Several potential methods of detecting loss-of-lock are suggested. They vary in the approach used, the assumptions made and the signals to be tested.

4.1 Tests For Randomness.

Theoretically, the measurement residual \tilde{z} of a properly functioning Kalman filter is supposed to be a zero mean, white random sequence with covariance equal to

$$\text{cov} [\tilde{z}(k+1|k)] = \text{HP}(k+1|k)\text{H}^T + \text{R}$$

The "Runs Test" of Appendix A can be applied to the measurement residuals to determine whether they are random or not. When the receiver is locked on so that valid range measurements are being received, \tilde{z} should be random. If the test indicates some non-random deterministic trend in the \tilde{z} then this would be attributed to loss-of-lock. Note that other modeling errors or filter simplifications might also cause \tilde{z} residuals to exhibit non-random behavior. Also, note that when the receiver bandwidth is narrow to guard against jamming, slow receiver dynamic response times will cause a deterministic transient component to exist in the \tilde{z} residuals. Whether or not this would invalidate the runs test conclusions remains to be evaluated.

In principle, the runs test could probably also be applied to the delay lock loop tracking error as well. This statement is based upon the supposition that the error in a properly functioning tracking loop will be randomly fluctuating about zero.

The advantage of a runs test is its extreme simplicity. It does not require knowledge of the signal variance. However, a fairly large number of samples (25 to 30 or more) must be tested in order that the

asymptotic gaussian properties be valid. These facts are what prompt the suggestion of using the delay lock loop tracking error signal, which is available at a much higher rate, and whose variance is probably unknown.

4.2 Simple Tests of Magnitude

When the theoretical values for signal variances are available, reasonableness tests on the magnitude can be made. This is currently done on the measurement residuals using a six-sigma threshold. Sigma is determined from the Kalman filter covariance results. Something of this sort may be combined with a runs test, as suggested at the end of Appendix A. However, a single sample test on each individual residual is not as statistically meaningful as a test based on a set of successive samples. Some tests which depend on a sequence of samples are given in the following sections.

Before concluding this section on simple magnitude tests, it is pointed out that once loss-of-lock has occurred, the variance of the measurement residual grows in a linear fashion with time (see Section 3.2). The slope of the growth is the random walk variance σ_n^2 , which would have to be estimated from tests of the loss-of-lock behavior. If this opening of the threshold is properly incorporated, then the rejection of valid pseudo-range data at later times, as discussed in Section 2.3, should be avoidable.

4.3 Tests For Distribution Parameters

The form of the pseudo-range signal, and its statistical characteristics, are known when the receiver is locked-on. By hypothesizing a random walk model for the receiver output after loss-of-lock, a second set of possible signal characteristics are derived. Appendix B gives

details of a sequential test on receiver output signals. Two thresholds are specified, depending on probability of miss and probability of false alarm. Then, each time a receiver signal is tested, three possible decisions exist; (1) receiver is locked on; (2) receiver had lost lock; (3) no definite decision can be made yet, so additional data should continue to be processed.

This test was originally set up to be applied to receiver outputs z , rather than measurement residuals \tilde{z} (or some other signal like Δz). The reason was that z is available every second, whereas \tilde{z} is only available every five seconds. However, when the details of Appendix B are examined, it is seen that the propagated covariance matrix $P(k+1|k)$ and the linearized measurement matrix, as well as the residuals, are really required at each signal test time. Therefore, although the complete Kalman update is not required every second, readout of several key Kalman filter quantities would be required.

The advantage of this sequential procedure is that it is able to detect trends that are developing, since it works on a growing sequence of data. The major disadvantage is that, as formulated here, the test is set up to decide whether loss-of-lock has occurred at a known, specified time. A non-sequential simplified version of the same test is described in Appendix C. The intention is to test, and make a decision about loss-of-lock, every five seconds. This is the Kalman filter cycle time. Data used in the decision would be the five receiver outputs (one per second) since the last Kalman update. The performances of these schemes when only five signals are used in the decision, will have to be evaluated. This is especially true in view of the fact that these signals will be time-correlated.

Exactly analogous procedures could be developed and applied to other signals, such as Δz , by modifying the appropriate expressions for means and covariances. Similar procedures have recently been proposed [6] for target detection, i.e., discrimination between warheads and decoys.

4.4 Test on the Variance of the Δz Signal

It has been suggested in Section 2.2 that the integrity of an independent IMU-derived estimate be maintained so that GPS signals can be compared with it. The difference, called Δz , is defined in Section 3.1 and the mean and covariance of this signal are given in Section 3.2. It is shown there that the difference between the covariance of a valid Δz (receiver locked-on) and a bad Δz (loss-of-lock has occurred) is a linear, growing function of time. Such a clear divergence of signal characteristics should be easily detectible. A Chi-Square test is presented in Appendix D for this purpose. To allow earliest possible decisions, it is suggested that the sample variance in Δz be computed for five signals over a five second period. This sample variance is then compared with a threshold which is a constant (which depends on the desired confidence level) times the pseudo-range variance. This test is almost as simple as the existing magnitude reasonableness test, and should be investigated for its effectiveness. Perhaps a combination of this type of test and the runs test of Section 4.1 will prove to be simple, timely and effective.

5.0 Results

Some preliminary numerical results have been obtained, using two sources of data. First, the delay lock loop error is considered, and then the Kalman filter residuals.

5.1 A Preliminary Look at Delay Lock Loop Errors

The delay lock loop error signal was obtained, in the form of strip chart recordings, for several cases. Four sets of error signals were obtained by manually sampling some of these curves. The resolution was not good and sampling quantization errors are probably large as a result. In the first two cases, the signal-to-noise ratio was $C/N_0 = 15.7$ db-Hz. The system was in the non-coherent mode, using the P-code, and there was no jamming. Error samples were read off the curves at 3 second intervals (the finest subdivision on the plot paper), and the errors were read to the nearest meter. It is known that in both these first cases, loss-of-lock did not occur.

When the runs test described in Appendix A was applied, the following results were obtained.

CASE	No. of Samples	Sample Mean Not Subtracted		Sample Mean Subtracted	
		No. of Runs	Test Statistic	No. of Runs	Test Statistic
1	25	6	-2.90	9	-.727
2	22	3	-2.42	8	-.764

It is seen that when the sample mean \bar{s} is not subtracted off, a smaller number of runs is obtained, and the hypothesis H_0 , (the samples are not from a zero mean random sequence) is selected at the .05 level in both cases. The means were $\bar{s} = -.74m$ and $3.77m$, respectively. In the first case, enough samples were near enough to zero to cause the shift in the number of runs, even for a rather small value of \bar{s} . In both cases, subtracting out the sample mean before applying the runs test caused the hypothesis H_1 to be selected.

When the Chi-square test of Appendix D was applied to the same two sets of data, the sample variances were determined to be $S^2 = 14.815m^2$ and $25.0m^2$, respectively. The value to be used for the parent population variances is not certain. Collins uses $R = 25m^2$ in the coherent mode and increases that to $75m^2$ in the non-coherent mode. Even when the smaller value is used, the critical value of $\frac{\chi^2_R}{N} = 45$ is far larger than S^2 so hypothesis H_1 (valid tracking data) would be selected in both cases.

Cases 3 and 4 were again non-coherent P-code cases, each with $C/N_0 = 8$ and with an accelerating user. In both cases it is known that lock was lost (delay error larger than 1 1/2 chips). Because of the different strip chart time scale, samples were taken only every 6 seconds. Results from these two cases are summarized below:

	Case 3	Case 4
No. of Samples	21	33
mean \bar{s}	3.84	-4.23
variance S^2	53.65	36.0
no. of runs without subtracting of mean	3	4
no. of runs after sub- tracting off the means	3	4
μ_U	11.29	17.36
σ_U	2.185	2.80
Test Statistic with mean subtracted	-3.794	-4.766

In these two cases the number of runs is not changed by subtracting out the mean (but the values for n_1 and n_2 were). In both cases 3 and 4, the test statistic z is so large as to make it exceedingly unlikely that these samples came from a random population. Thus, H_0 is accepted. This is the correct decision.

With regard to the Chi-square test, the critical threshold at the 1% level is $1.79R$. This is to be compared with S^2 in order to select either H_0 or H_1 . The values $S^2 = 53.65m^2$ and $36.0m^2$ were found for cases 3 and 4. Naturally, the value used for R is crucial. If $R=75m^2$ is used, then the result is the selection of H_1 (a wrong decision). If $R=25m^2$ is used, the correct decision H_0 is made in case 3. In case 4, the Chi-square test does not select the correct hypothesis H_0 . It is noted that the current 6-sigma reasonableness test would not have thrown out any of the data points in case 3 and 4 (even assuming $R=25m^2$ is used).

Although intended for use with Kalman filter residuals, the maximum likelihood ratio tests of Appendices B and C can be applied to the delay error if a few assumptions are made to fill in for some lacking information. The delay error samples are used as if they are measurement residuals $\tilde{z}(k)=z(k)-H\hat{x}(k|k-1)$ of Eq (B-13). The term $z(k)-H\hat{x}(0|-1)$ can be approximately by $\tilde{z}(0)+\dot{z}\Delta t$. A value for $HPH^T + R$ of $50m^2$ is used. Finally, values for σ_n^2 and \dot{z} must be assumed, as well as values for P_M and P_F .

In cases 1 and 2, with no user motion, a value of $\dot{z}=0$ was used. If $P_M=P_F=.01$, then the two decision thresholds are -4.595 and 4.595 . If $P_M=P_F=.05$, they reduce to -2.944 and 2.944 . When $\sigma_n^2=.3(m/sec)^2$, both cases 1 and 2 gave a value of Λ which exceeded the upper threshold of 2.944 but not 4.595 . This means that decision H_1 is reached at the .05 level, but no decision was reached at the .01 level. Increasing σ_n^2 or decreasing HPH^T+R caused the correct decision to be made with greater certainty after fewer samples. When the same procedure was applied to cases 3 and 4, results were very sensitive to the assumed value for \dot{z} as well as σ_n^2 and HPH^T+R .

Because of the coarse quantization of the delay error signals, and because so many assumptions had to be made to fill in for unknown parameter values, the preceding results are quite tenuous. While no strong final conclusions can be made about the suggested methods, the results do seem hopeful.

5.2 A Preliminary Look at Kalman Filter Residuals.

A digital tape of some Kalman filter-related results from Acceptance Test Plan 48 was received from Collins Radio during the final week of this study. This is the source of the data used here.

The period of time from 4863 sec. (Nav time) to 5043 sec. will be examined. During this period several interesting events occur. At 4888 sec. the range residual in Channel 1 suddenly exceeds the 6-sigma test and is, therefore, not used by the filter for the succeeding period of time. At 4978 sec. the residual of Channel 2 also exceeds the 6-sigma limit and is thereafter ignored by the filter. At 5038 sec. the residual of Channel 1 drops back below the 6-sigma threshold and thereafter stays far within that bound. Channel 2 stays outside the bound. During this period, Channels 3 and 4 have small residuals that never come close to their 6-sigma bounds.

The system is operating in the non-coherent mode. The user is stationary and the nominal signal strength is -169db. There is 61db of jamming power during the entire time selected above for analysis.

Within the above time range, 35 data points were available for each channel. One sample is given every five seconds, except that the data for 4943 sec. and 4958 sec. are missing for some unknown reason.

The first obvious impression from the residual data is that none of the four channels look like zero mean random processes that a Kalman filter should theoretically produce. In fact, Channel 1 has only 2 sign changes (3 runs) out of 35 points. Channel 2 has no sign changes (1 run), Channel 3 has 4 sign changes (5 runs) and Channel 4 has 1 sign change (2 runs). There is no point doing the complete runs test computation with this type of data. Based on results of Section 5.1 it is felt that the sample mean should always be subtracted at first. For Channel 4, this mean is $\bar{s} = -5.1287m$, and the sample variance is $S^2 = 41.77m^2$. Using $s - \bar{s}$, it is found that Channel 4 has 4 transitions ($U=5$ runs), and $\mu_U = 18.486$ and $\sigma_U = 2.91$. The test statistic is $z = -4.63$.

This means that if these residuals did come from a random process of mean \bar{s} , then a very, very unusual 4.63 sigma event has occurred. In fact, Channel 4 residuals are not random, but systematic. That is, hypothesis H_0 of Appendix A is selected. Deciding that loss-of-lock has occurred is an incorrect decision. There are other causes for the non-random residual behavior. Since the magnitudes of all Channel 4 residuals are quite small, this situation corresponds to occurrence 3 of Figure A-4 (small error, but not random).

An analysis of Channel 3 shows much the same thing. That is, $\bar{s} = -1.2504\text{m}$ and $S^2 = 15.382\text{m}^2$. Then $(s - \bar{s})$ has 5 runs, and coincidentally, the same results as Channel 4, $n_1 = 17$, $n_2 = 18$. Thus, the residuals of Channel 3 are also small, but not random.

The digital test tape also lists the theoretical variance $\text{HPH}^T + R$ from the Kalman filter for each channel. For Channels 3 and 4, this theoretical variance is almost constant during the time of interest. Values range from 87 to 92m^2 . Of this total, 75m^2 is due to the input value for R used by Collins in the non-coherent mode. Two conclusions can be drawn from this: (1) any test (such as the Chi-square test) which depends on a larger than theoretical variance to signal loss-of-lock is not likely to work very well because the sample variance S^2 is so much smaller than the theoretical values. That is, the residuals really do not have the statistical properties that were assumed for them when setting up the mathematical models for hypothesis testing. An R value of 75 is far larger than reality; (2) with $R = 75$, it follows that HPH^T is on the order of 12 to 17m^2 . It can be concluded that on a magnitude basis, the Kalman gain for these channels is about .14 to .18. This means that the filter update is making very small corrections to the state estimates. Rather, 82 to 86% of the weight of the updated state estimate is placed on the propagated value of the last estimate. Only 14 to 18% of the weight is placed on information gained from the GPS measurement.

A closer look at Channels 1 and 2 is in order. They will also fail the randomness test (runs test) but this is not very significant in view of the fact that even the "good" channels fail this test. The random walk model of loss-of-lock is clearly not applicable to the data of this case. The reason is that in both cases where the residual exceeds its 6-sigma limit (suddenly) it is caused by the deterministically computed "ionospheric correction". For example, at 4883 sec. the Channel 1 range residual is -0.193m and the ionospheric correction is 17.872m. At 4888 sec. (time that residual first exceeds 6-sigma) the residual jumps to 60.193m and the ionospheric correction is 77.963m, or 60.151m larger than the previous correction. The correction prior to 4883 sec. had also been exactly 17.872m at least as far back as 4753 sec.

A similar jump in the Channel 2 ionospheric correction (from 14.491m to 77.286m) at 4978 sec. accounts for the range residual jump from 12.282m to 74.604m. The jump is 62.795m in ionospheric correction and 62.322m in range residual.

Clearly, the "corrections" caused the so-called loss-of-lock as judged by the 6-sigma reasonableness test. This, plus the fact that the filter residuals are not random, do not have zero means and have variances far smaller than the filter is told to expect, makes further testing of this data of academic interest at best.

The reason for the non-randomness of residuals is the long receiver time constant when faced with jamming in the non-coherent mode. Measurement errors, and, hence, filter residuals, are strongly time correlated with correlation times on the same order of magnitude as receiver time constants (20 to 30 seconds). This correlation is not modeled into the filter. Instead, the value of the measurement noise variance R is increased from 25m^2 to 75m^2 to de-weight the measurements and as an ad hoc means of coping with correlations. This increase in R is also the reason why sample variances (S^2 of 15 to 42m^2) are so much smaller than the filter predicts (σ^2 of 87 to 92m^2).

In summary, non-random residuals do not necessarily indicate loss-of-lock, but rather suboptimal filter modeling. Small residual variances do not guarantee good tracking but may indicate an excessively large standard of comparison due to non-optimal filter parameters.

6.0 Conclusions and Recommendations

Methods of detecting loss-of-lock in the Generalized Development Model of the GPS receiver have been developed. Some tests of the methods on available data have been presented. Promising or hopeful results were obtained when the delay lock loop error signal was used, although only a few test cases were analyzed.

When the methods were tested on Kalman filter residuals, the results were not very satisfactory because the filter is of suboptimal design. As a result, the residuals bear little resemblance to their theoretical model. This leads to results which are contradictory to expectations. In Section 2.3 it was stated that filter covariances would be overly optimistic. After seeing the data, late in this study, it is clear that at least the residual covariances are far too pessimistic. Unmodeled receiver states, and the resulting correlated measurement errors are probably the major sources of trouble. In some ways the ad hoc increase in noise covariance R compounds the problem of using residuals to detect loss-of-lock, since this artificially causes a great discrepancy between actual and theoretical residual variances.

The reason for increasing R is to de-weight measurements somewhat in the non-coherent mode. It could be that they are de-weighted too much, to the point where the GPS signals and the Kalman filter are having a very minor effect on state estimates.

It has been determined, at least in the few cases reported here, that "loss-of-lock" as defined by a filter residual exceeding 6-sigma, is actually caused by a sudden jump in the ionospheric correction term.

A reasonableness check on these corrections must be added. This should be very easy to do. The ionospheric corrections remain constant over relatively long periods, and any sudden drastic jump should be edited out. In fact, this correction term may have value as an indicator of jamming.

The recommendation of Section 2.2 is reiterated. The feedback to the IMU should be disabled at the time of a switch over to the non-coherent mode. The measurement difference signal Δz defined in Section 3.1 should be investigated for its suitability in detecting loss-of-lock. The methods of this report are suggested. Lack of suitable data prevented the Δz signal from being tested here.

A combination of methods, such as a test for randomness and a test on magnitude or a Chi-square test on variances may be the best procedure.

A final suggestion for further consideration is that the time average approximation for the residual variance, called S^2 in the report, could prove useful in a simple adaptive method of adjusting the value of R in real time.

REFERENCES

1. W. L. Brogan and A. C. Liang, Navigation of a Tactical Aircraft Using a Fully Integrated DNSS/Strapped-Down Inertial System, TOR-0073 (3020-03)-2, The Aerospace Corp. El Segundo, CA, March 1973.
2. AFAL Generalized Development Model of GPS User Equipment, (Vol. 1, Hardware, Vol. 2, Software) Collins Radio Group/Rockwell International, Aug 1975.
3. P. H. Yeh, A Treatise on Anti-jamming Margin of an IMU/Computer Aided Global Positioning Navigation System, TOR-0076(6473-01)-1, The Aerospace Corp. El Segundo, CA, Oct. 1975.
4. S. F. Russell, Loss-of-Lock Detection Summary, Internal Letter No. GDM-227, Rockwell International, 22 June 1978.
5. G. A. Madrid and G. J. Bierman, "Application of Kalman Filtering to Spacecraft Range Residual Prediction", IEEE Transactions on Automatic Control, Vol. AC-23, No. 3, June 1978, pp. 430-433.
6. C. W. Therrien, "A Sequential Approach to Target Discrimination", IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-14, No. 3, May 78, pp. 433-440.
7. K. A. Brownlee, Statistical Theory and Methodology in Science and Engineering, John Wiley and Sons, pp. 164-170.
8. A. P. Sage and J. L. Melsa, Estimation Theory with Applications to Communications and Control, McGraw-Hill Book Co., Inc., 1971.
9. P. G. Hoel, Introduction to Mathematical Statistics, John Wiley and Sons, 1962.

Appendix A

Runs Test [7]

In order to test the hypothesis that a given sequence of signal samples is random, as opposed to having some deterministic trend in it, a "runs test" can be used, as follows. Consider a sequence of signals $s(k)$ for $k=1, 2, \dots, N$. The hypotheses are:

H_0 : $s(k)$ is not a zero mean, random sequence

H_1 : $s(k)$ is a zero mean, random sequence.

(For a known, non-zero mean $m(k)$, a test for randomness can still be applied by using $s^1(k) = s(k) - m(k)$, that is by constructing a zero mean process first.) Heuristic motivation: If $s(k)$ satisfies H_1 , then one would expect the samples to change sign in some random pattern. Too many sign changes, like every sample or two, might indicate that $\{s(k)\}$ is a high frequency deterministic oscillation. Too few sign changes might indicate another type of deterministic trend, e.g., just one sign reversal might indicate a ramp or parabolic signal behavior.

Definition: A "run" consists of one or more consecutive samples of the same sign. The number of runs in a string of N test samples will be denoted as U , and under H_1 , U will be random variable. The test statistic U has a density function which asymptotically approaches the gaussian density as the size of the sample, N , increases. In practice, if N exceeds 25 or 30, the gaussian approximation may justifiably be used. This assumption will be used here. The mean and standard deviation of the random variable U are [7]

$$\mu_U = \frac{2n_1n_2}{N} + 1$$

and

$$\sigma_U = \left[\frac{2n_1n_2(2n_1n_2 - N)}{N^2(N-1)} \right]^{1/2}$$

where

n_1 = number of times $s(k)$ is positive
in N samples

n_2 = number of times $s(k)$ is negative
in N samples

$$N = n_1 + n_2$$

A sketch of the density function $f(U)$, assuming H_1 is true is shown below

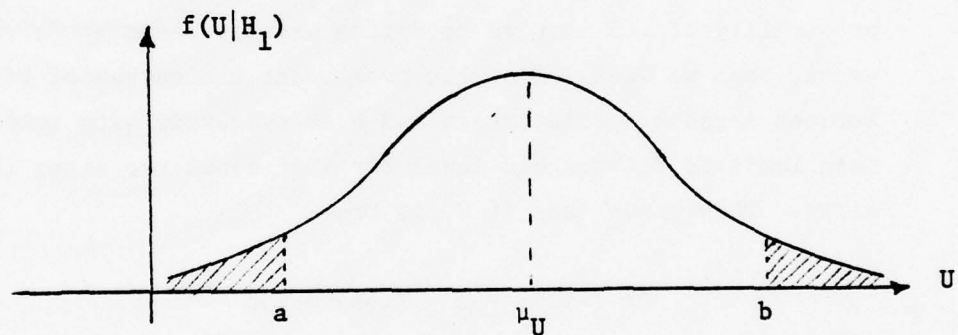


Fig. A-1 Density Function for the
Test Statistic U

If H_1 is indeed true, then on a given test of N samples, the number of runs U would be expected to fall between the limits $[a, b]$ with a probability which can easily be computed for specified values of a and b . If U exceeds b or is less than a , then either

- (1) H_1 does not hold and, therefore, H_0 is chosen, or
- (2) A very rare probabilistic event has occurred.

In practice, one would carry out the following procedures:

- (1) Select N samples and observe n_1 , n_2 , and U .
- (2) Compute μ_U and σ_U .
- (3) Normalize to allow use of zero mean, unit variance normalized tables, i.e., $z = (U - \mu_U)/\sigma_U$ becomes the test statistic.
- (4) Select the desired confidence level. This establishes how large the areas are in the tails of the curve of Figure A-1, which establishes the rejection threshold for the test. For example, if a probability of .05 is used to define what is meant by "a very rare" event, then we want 95% of the area under the curves of Figure 1 to be between a and b . Selecting a and b as symmetric with respect to the mean leads to a threshold level for z of about two sigma (1.96 actually). This means that if H_1 is true,

$$\Pr (-1.96 \leq z \leq 1.96) = .95$$

Thus, if the computed value for z in item 3 satisfies $-1.96 \leq z \leq 1.96$, then there is no reason to reject H_1 and accept H_0 . On the other hand, if z is outside this range, the conclusion is that H_0 applies, i.e., $\{s(k)\}$ is not random. We accept the chance of being wrong 5% of the time.

A possible implementation: If the samples $s(k)$ are encoded as 1 if $s(k) \geq 0$ and 0 otherwise, then an N -bit shift register could be used to maintain the most recent N samples on which the test for randomness will

be applied. The algebraic sum of the bits in this shift register is n_1 . $N - n_1$, gives n_2 , and the value of U can be determined by counting the number of transitions from 0 to 1 and from 1 to 0. The number of runs U and the number of transitions t are related by

$$U = t + 1$$

Thus, all quantities needed for the runs test are easily obtained.

As an illustrative example, the velocity residuals from a two-state extended Kalman filter simulation with range-rate measurements are used as the test signal $s(k)$

k	s(k)	
0	-.02	Run #1
.5	-3.34	
1.0	-.49	
1.5	-2.91	
2.0	-2.67	
2.5	-.80	
3.0	-.67	
3.5	.65	2
4.0	-.76	3
4.5	.39	4
5.0	-1.67	5
5.5	-.51	
6.0	.99	6
6.5	3.6	
7.0	2.77	
7.5	1.61	
8.0	-.15	7
8.5	-.62	
9.0	-1.15	
9.5	-1.86	
10.0	-.96	

$U = 7$ runs, transitions = 6

$N = 21$ samples

$n_1 = 6, n_2 = 15$

Table A-I Some Example Statistics, Showing the Runs

The question is, are these residuals random (i.e., is the filter working properly) or is there a systematic trend, due to a modeling error or some other cause?

The runs test, at the 5% level, is applied.

$$\mu_U = \frac{2(6)(15)}{21} + 1 \approx 9.57$$

$$\sigma_U = \sqrt{3.24} \approx 1.8$$

\therefore the test statistic is

$$z = (U - \mu_U) / \sigma_U \approx -1.4$$

Since $-1.96 \leq z \leq 1.96$, there is no justification for saying that this sequence is non-random (even though there are several long strings of consecutive negative residuals).

Note that the runs test does not require any knowledge of the statistics of the signal samples $s(k)$ beyond the assumption of zero mean. Therefore, this test could be applied to any receiver signal which is supposed to be random when the code is being tracked; for example, the delay error in the delay lock loop. The advantage of using this signal as opposed to the Kalman filter pseudo-range residuals is the availability of data at a higher rate. Thus, N samples can be tested in a relatively shorter time.

Disadvantages associated with the "runs test" approach to detecting loss-of-lock are:

- (1) It may indicate that the test signal is non-random because of a relatively small and tolerable system bias as sketched in Figure A-2.

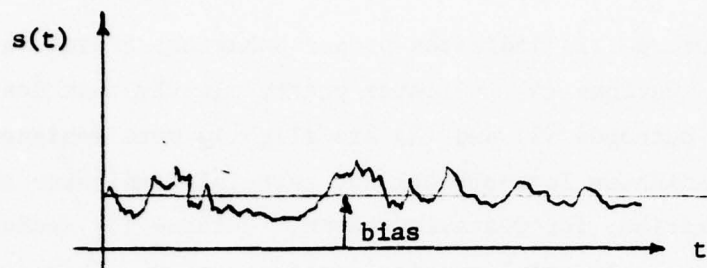


Fig. A-2 A Biased Signal

(2) It may indicate that a wildly erratic signal is indeed random, although, in fact, the error levels are so large as to preclude any useful processing of the signal. Figure A-3 represents such a sample.

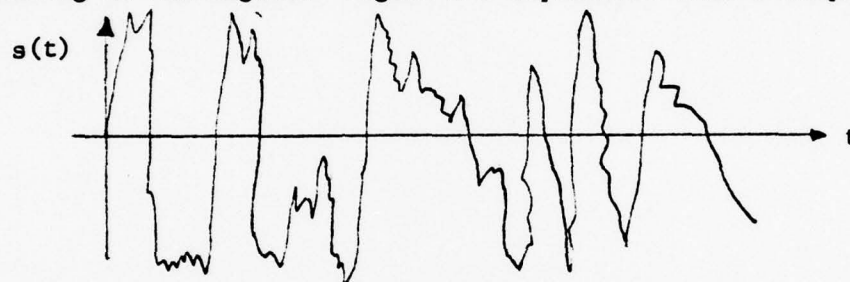


Fig. A-3: A Random Signal With Wild Fluctuations

Both of these possible defects could be overcome by combining a magnitude threshold test with the runs test. Figure A-4 shows how such a combination might be used:

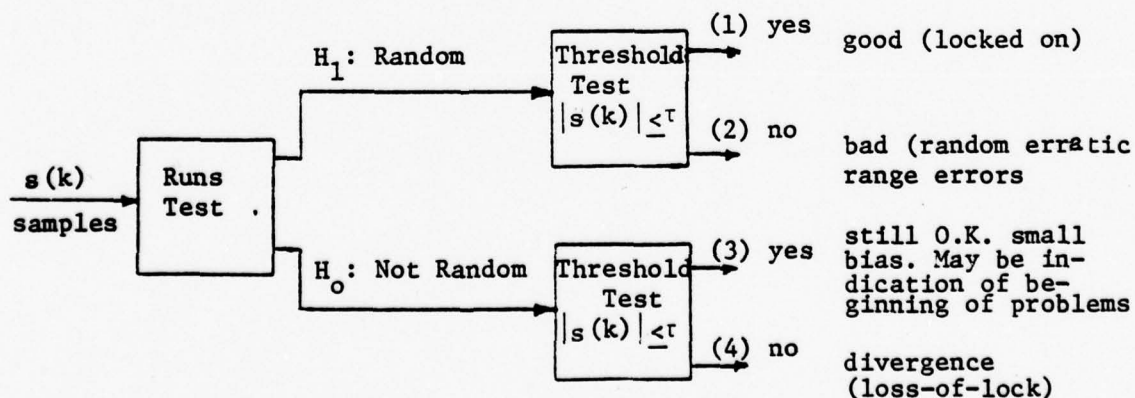


Fig. A-4: Combined Runs Test and Threshold Test

Possible outcome (1) indicates proper behavior; therefore, the system is locked-on. Outcome (4) indicates pretty clearly that loss-of-lock has occurred. Outcomes (2) and (3) are slightly more ambiguous, but (2) probably indicates loss-of-lock and certainly indicates unacceptable system operation, for whatever reason. Outcome (3) indicates that the system is probably still working satisfactorily. This could mean there is some small system bias, but it might also indicate the beginning of a slow divergence and the situation should continue to be carefully monitored.

Appendix B*

Sequential Likelihood Ratio Test for Detecting Loss-of-Lock

The problems being considered is the determination of whether the measurement (receiver output) is a valid pseudo-range measurement or some sort of random walk because loss-of-lock has occurred. That is, one of two alternative hypothesis must be selected.

$$\begin{aligned} H_0: z &= w + v \\ \text{or} & \\ H_1: z &= \underbrace{r + ct_b}_{\triangleq r_p} + v = H\underline{x} + v \end{aligned} \quad (B-1)$$

where v is random, white zero mean noise with variance R and w is a random walk variable,

$$w = \int_0^t n(\tau) d\tau + r_p(o) \quad (B-2)$$

The noise n is also white, zero mean with variance σ_n^2 (latter assumed constant) and $r_p(o)$ is the pseudo-range which existed at the time of loss-of-lock. Also, t_b is the time bias, c is the speed of light, \underline{x} is the system state vector and H is the linearized measurement matrix. If z is assumed gaussian, then its density function is determined by its first two moments. First consider the situation when H_0 is true. Then

$$E\{z\} = E\{w\} + E\{v\} = \int_0^t E\{n\} d\tau + E\{r_p(o)\} + E\{v\} \quad (B-3)$$

*In this appendix, a linear relationship between state and measurement is used only for simplicity of exposition.

Since n and v have zero means,

$$E\{z\} = E\{r_p(o)\} = H(0) \underline{\bar{x}}(0)$$

Also, $\text{cov}\{z\} = E\{[z - \bar{z}]^2\}$, but

$$z - \bar{z} = \int_0^t n(\tau) d\tau + v + H(o) [\underline{x}(o) - \underline{\bar{x}}(o)]$$

so that under hypothesis H_0

$$\text{cov}(z) = H(o) \text{cov}[\underline{x}(o)] H^T(o) + R + \sigma_n^2 t \quad (B-4)$$

The time o is being used to indicate the point where loss-of-lock occurred, and not some initial problem time. Also, note that unconditioned expectations have been used above, so that $\text{cov}[\underline{x}(o)]$ is not $P(o|o)$ or $P(o|-1)$, the conditioned covariances of the Kalman filter. However, we will ultimately switch to a recursive formulation and then it will be seen that quantities from the Kalman filter will become involved.

Now consider H_1 to be true. Then

$$E\{z\} = E\{r_p(t)\} = H\underline{\bar{x}}(t) \quad (B-5)$$

and

$$\text{cov}\{z\} = H(t) \text{cov}[\underline{x}(t)] H^T(t) + R \quad (B-6)$$

The two possibilities H_0 and H_1 lead to two separate density functions $f(z|H_0)$ and $f(z|H_1)$. Under the gaussian assumption, the above means and covariances suffice to completely specify these density functions.

Define the likelihood ratio as

$$L^1 \triangleq \frac{f(z|H_1)}{f(z|H_0)} \quad (B-7)$$

Consider the densities sketched in Fig. B-1, and assume that a particular observation z^1 is obtained from the receiver.

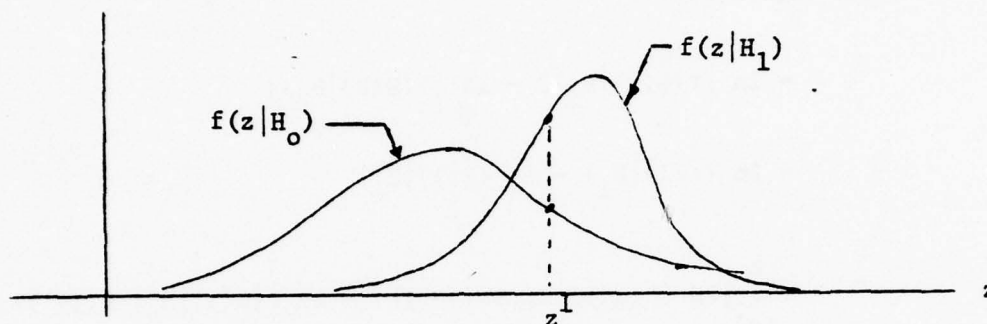


Figure B-1: The Two Possible Density Functions

Clearly at z^1 we have $L^1 > 1$ and a maximum likelihood decision rule would select alternative H_1 in this case.

Actually, the decision is to be based on a sequence of measurements, $\mathbf{z}(k) = \{z(1), z(2), \dots, z(k)\}$ rather than on a single measurement $z(k)$. Thus, a modified likelihood ratio is defined as

$$L \triangleq \frac{f(\mathbf{z}|\mathbf{H}_1)}{f(\mathbf{z}|\mathbf{H}_0)} \quad (\text{B-8})$$

Note that $f(\mathbf{z}(k)|\mathbf{H}_1) = f(z(k)|\mathbf{H}_1, \mathbf{z}(k-1))f(\mathbf{z}(k-1)|\mathbf{H}_1)$. This can be repeatedly factored, leading to

$$\begin{aligned} f(\mathbf{z}(k)|\mathbf{H}_1) &= f(z(k)|\mathbf{H}_1, \mathbf{z}(k-1))f(z(k-1)|\mathbf{H}_1, \mathbf{z}(k-2)) \dots * \\ &\quad * f(z(2)|\mathbf{H}_1, \mathbf{z}(1))f(z(1)|\mathbf{H}_1) \\ &= f(z(1)|\mathbf{H}_1) \prod_{j=2}^k f(z(j)|\mathbf{H}_1, \mathbf{z}(j-1)) \end{aligned} \quad (\text{B-9})$$

The same form holds for both $i=0$ (H_0) and $i=1$ (H_1).

The natural log of $L(k)$ is defined as $\Lambda(k)$, and can be written as

$$\Lambda(k) = \ln\{L(k)\}$$

$$= \ln [f(z(k)|H_1)] - \ln [f(z(k)|H_0)]$$

$$= \ln (z(1)|H_1) - \ln (z(1)|H_0)$$

$$+ \sum_{j=2}^b [\ln \{f(z(j)|H_1, z(j-1))\} - \ln \{f(z(j)|H_0, z(j-1))\}]$$

$$= \Lambda(k-1) + \ln \{f(z(k)|H_1, z(k-1))\}$$

$$- \ln \{f(z(k)|H_0, z(k-1))\} \quad (B-10)$$

If the original density functions are gaussian, then so are all the densities conditioned on past measurements. The results of Eq (B-3) through Eq (B-6) must be modified slightly to obtain the moments for $f(z(k)|H_1, z(k-1))$ for example. The modification consists of using $E\{\cdot | z(k-1)\}$ instead of the unconditioned expectations used earlier. Thus, the former \bar{x} terms become $E\{\underline{x}(k) | z(k-1)\}$, which is the usual Kalman estimate $\hat{x}(k|k-1)$. Also, $\text{cov}[x]$ terms become $\text{cov}[\underline{x}(k) | z(k-1)]$ which is the usual Kalman filter covariance $P(k|k-1)$. These manipulations lead to

$$f(z(k)|H_1, z(k-1)) = \frac{\exp \{-1/2 [z(k) - H\hat{x}(k|k-1)]^T [HPH^T + R]^{-1} [z(k) - H\hat{x}(k|k-1)]\}}{(2\pi)^{m/2} |HP(k|k-1)H^T + R|^{1/2}}$$

Since each receiver channel is considered separately, $m=1$, z is a scalar and the above expression simplifies to

$$f(z(k)|H_1, z(k-1)) = \frac{\exp \{-1/2 [z - H\hat{x}(k|k-1)]^2 / [HP(k|k-1)H^T + R]\}}{\sqrt{2\pi} [HP(k|k-1)H^T + R]^{1/2}} \quad (B-11)$$

When the same kinds of manipulations are applied to $f(z(k)|H_0, z(k-1))$, the result is

$$\begin{aligned} & f(z(k)|H_0, z(k-1)) \\ &= \frac{\exp \{-1/2 [z(k) - H\hat{x}(o|-1)]^2 / [H(o)P(o|-1)H^T(o) + R + \sigma_n^2(t_k - t_o)]\}}{\sqrt{2\pi} \{H(o)P(o|-1)H^T(o) + R + \sigma_n^2(t_k - t_o)\}^{1/2}} \quad (B-12) \end{aligned}$$

Note: t_o is used to indicate the time loss-of-lock occurs, and not some initial time. It could be any time in the range

$$t_o \in [t_1, t_k]$$

Taking natural logs gives the log-likelihood ratio in a recursive form

$$\begin{aligned} \Lambda(k) = & \Lambda(k-1) + 1/2 \ln \{H(o)P(o|-1)H^T(o) + R + \sigma_n^2(t_k - t_o)\} \\ & - 1/2 \ln \{H(k)P(k|k-1)H^T(k) + R\} \\ & + 1/2 [z(k) - H\hat{x}(o|-1)]^2 / [H(o)P(o|-1)H^T(o) + R + \sigma_n^2(t_k - t_o)] \\ & - 1/2 [z(k) - H\hat{x}(k|k-1)]^2 / [H(k)P(k|k-1)H^T(k) + R] \quad (B-13) \end{aligned}$$

An initial value $\Lambda=0$ starts the process. This says that initially the presence or absence of a valid pseudo-range signal is equally likely. The quantities $P(k|k-1)$, $z(k)$ and $\hat{x}(k|k-1)$ are the usual current values of covariance, measurement and state estimate from the Kalman filter. The same quantities with a zero time argument in place of k , apply at the time that loss-of-lock is presumed to have occurred.

In order to use $\Lambda(k)$ in a sequential hypothesis test [8], two thresholds τ_0 and τ_1 must be specified. Then;

$$\begin{aligned} &\text{if } \Lambda(k) \leq \ln(\tau_0) \text{ accept } H_0 \\ &\text{if } \Lambda(k) \geq \ln(\tau_1) \text{ accept } H_1 \\ &\text{if } \ln(\tau_0) < \Lambda(k) < \ln(\tau_1) \text{ continue testing,} \\ &\quad \text{no decision yet.} \end{aligned} \tag{B-14}$$

The decision thresholds are selected as follows: Define P_M as the probability of miss (P_M = Probability that hypothesis H_0 is accepted when, in fact, H_1 is true; i.e., we miss the fact that a valid range measurement is present). Also define a probability of false alarm P_F (P_F = Probability that H_1 is accepted when in fact H_0 is true; i.e. it is falsely concluded that a valid range signal is present). Acceptable numerical values must be specified for P_M and P_F . These are then used to find the necessary thresholds

$$\tau_0 = \frac{P_M}{1-P_F} \quad \text{and} \quad \tau_1 = \frac{1-P_M}{P_F} \tag{B-15}$$

Figure B-2 illustrates the use of the sequential likelihood ratio test,

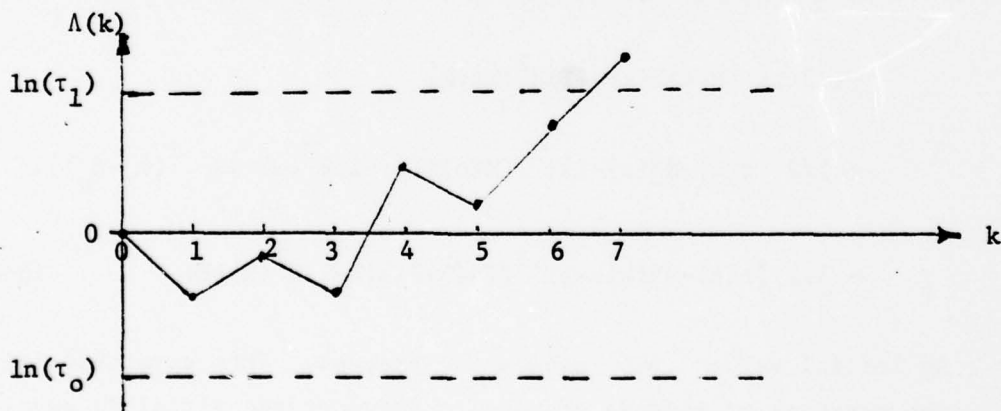


Figure B-2 A Possible Sequence of $\Lambda(k)$ Values

For the situation shown in Fig. B-2, the two alternate hypotheses are that loss-of-lock has occurred at time 0 and that loss-of-lock has not occurred. The first six samples yield no conclusion, since neither threshold is crossed. Only at sample 7 is a definite decision made, and in this case it is that H_1 is true, loss-of-lock has not occurred. On another set of sample measurements, $\Lambda(k)$ might cross the lower threshold, in which case H_0 would be accepted; that is, it would be decided that loss-of-lock did occur at time 0.

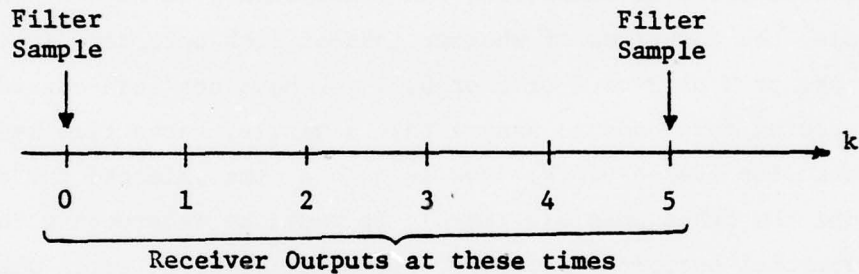
In the GPS receiver problem, complicating issues arise regarding the above procedure. First, two simple alternatives do not really describe the true situation. In relationship to Figure B-2, for example, the questions of whether loss-of-lock occurred, (not at $k=0$, but at $k=1$ or 2 or 3 or 4 or 5 or 6, . . . have not been considered. The preceding developments assume that a single, known time $k=0$ is being tested for loss-of-lock. How is such a time selected for test, and what about the other possible times? It would be interesting to test the sequential decision procedure, via simulation, in cases where the time loss-of-lock occur can be controlled. These simulated tests could prove or disprove the feasibility of the sequential likelihood ratio test and could shed light on how long it takes before loss-of-lock can be detected in this idealized case.

A more pragmatic approach, using Eq (B-13) of this appendix but in a non-recursive manner on a fixed set of data points, is suggested in Appendix C.

APPENDIX C

A Nonsequential Likelihood Ratio Test For the GPS Receiver

Each channel of the GPS receiver produces an output once per second. These are sampled and processed by the Kalman filter once every five seconds. Four out of each five receiver outputs are not used at present. It is suggested here that all receiver outputs be used in detecting loss-of-lock. Let $k=0$ represent a receiver output which is sampled for use in the filters. The next sample which the filter will process is $k=5$, as shown below.



In deciding whether or not $z(0)$ is a valid pseudo-range measurement, the five measurements $\mathbf{z} = \{z(0), z(1), z(2), z(3), z(4)\}$ can be used in a non-recursive, fixed data span mode as described below. Then, the whole process can be repeated using $\mathbf{z} = \{z(5), z(6), z(7), z(8), z(9)\}$ for the next filter cycle, and similarly for succeeding cycles. In this mode of operation, the two alternate hypothesis of Appendix B are valid and appropriate at each step (since a decision will be made about loss-of-lock for every Kalman filter input). Eq (B-13) for computing $\Lambda(k)$ is still valid, but now just five sets of measurements $z(k)$ will be used and a decision is only made once for each set of five. At the beginning of each set of five samples, is initialized to zero. In order to perform such a test, it is clear that $P(k|k-1)$ and $\hat{\mathbf{x}}(k|k-1)$ will be needed

at four times when they would not normally be used by the Kalman filter. Since no Kalman updates are done between $k=0$ and $k=5$, it is clear, for example, that $P(3|2) = P(3|0)$ and $\hat{x}(3|2) = \hat{x}(3|0)$. Thus, it appears that all needed quantities can be obtained by just interrupting the normal time propagation process at the appropriate times. No extra computations are required.

After accumulating Λ for a complete set of five measurements, the test for loss-of-lock is

$$\begin{array}{ccc} & \text{accept } H_1 & \\ \Lambda & \begin{array}{c} > \\ < \end{array} & 0 \\ & \text{accept } H_0 & \end{array} \quad (C-1)$$

Equation (C-1) is, of course, entirely equivalent to saying that H_1 is accepted if the likelihood ratio L of Eq (B-8) is greater than 1 and H_0 is accepted otherwise.

APPENDIX D

χ^2 Test on the Signal Δz .

Given N samples of Δz_i , compute the sample mean

$$\overline{\Delta z} = \frac{1}{N} \sum_{i=1}^N \Delta z \quad (D-1)$$

and the sample variance

$$S^2 = \frac{1}{N} \sum_{i=1}^N (\Delta z_i - \overline{\Delta z})^2 \quad (D-2)$$

Then, the test statistic

$$\zeta = \frac{NS^2}{R} \quad (D-3)$$

has a χ^2 distribution with $N-1$ degrees of freedom [9]. This distribution is sketched below

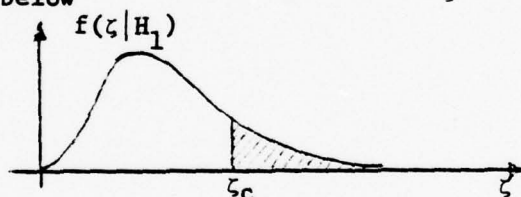


Fig. D-1: χ^2 Density Function

A value of ζ_c can be selected such that the cross-hatched area in the tail is the desired probability of rejecting H_1 when it is actually true. This probability is P_M of Appendix B. For example, if $N=5$, then $\zeta_c = 13.28$ for $P_M = .01$ and $\zeta_c = 9.49$ for $P_M = .05$. (These values loosely correspond to three-sigma and two-sigma levels of confidence in the gaussian case.)

For a particular sample value for ζ of Eq (D-3);

if $\zeta > \zeta_c$, the decision is to accept H_0

if $\zeta \leq \zeta_c$, the decision is to accept H_1

Numerically, if $N=5$ and $P_M=.01$ then $\zeta_c = 13.28$ so that the critical threshold for S^2 is $\zeta_c R/N = 2.66R \approx 3R$.

As a more concrete and complete example, consider the 21 samples listed in Table A-I. The sample mean is computed as -0.40762 and the sample variance is $S^2 = 2.81384$. Suppose a probability of miss $P_M=.01$ is assigned. Then from chi square tables, $\zeta_c=37.57$. Suppose, purely for the sake of this illustrative example, that the samples come from a population whose variance is $R=1$ (if loss-of-lock has not occurred, i.e., if H_1 is true). We thus compare $S^2 = 2.81384$ with $\zeta_c R/N = 37.57/21 \approx 1.79$. Since S^2 exceeds this threshold, we conclude with a probability of 0.99 that the population variance must be larger than $R=1$. This means we accept hypothesis H_0 that loss-of-lock has occurred.

Note that for this same set of hypothetical parameter values the existing six-sigma test would not have rejected any of the 21 samples. They all have magnitude less than $6R=6$. In fact, only two samples exceed a three-sigma threshold.

This sample was hypothetical because the value $R=1$ was selected only to make a point. The 21 samples are really Kalman filter residuals and their theoretical variance is of the form $HP(k+1)H^T + R$, which is a time-varying quantity. In the case which generated the samples of Table A-I, this residual variance was decreasing but had an average of about 20 to 25 over the time span of interest. If this value is used instead of the formerly assumed value of 1, hypothesis H_1 would be accepted.

Note that the χ^2 test is suggested for the signal Δz rather than the residual \hat{z} . This is because the variance of Δz does not depend on the Kalman covariance matrix P , but only on the measurement noise covariance R , which is constant. See Section 3.2.

Physics of Matrix Cathodes

T. P. Graham

Physics Dept.

University of Dayton

Final Report for Period

5 June 1978 to 11 August 1978

Work Performed While a USAF - ASEE

Summer Faculty Fellow

at the

Avionics Laboratory, WPAFB

Thomas P. Graham
Dept. of Physics
University of Dayton

ABSTRACT: Physics of Matrix Cathodes

Nickel matrix cathodes are currently under investigation with the expectation that they may be able to fulfill the emission requirements of microwave tubes in the near future. A facility for the measurement of the thermionic emission characteristics of test cathodes was assembled. This facility combines high vacuum instrumentation, mass spectrometric analysis, data logging, computation and display capabilities in a semi-automatic test and measurement system. The system is suitable for routine evaluation of test cathodes and life testing as well as for more detailed investigations into the physics of the activation and emission behavior of cathodes. Schematic diagrams and preliminary data for some cathodes are presented.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the financial support of The Ohio State University under whose auspices the USAF - ASEE Summer Faculty Program at Wright Patterson Air Force is administered. My thanks go also to the members of the Microwave Devices Group of the **Avionics** Laboratory who generously provided assistance to me in this project. In particular, I acknowledge the assistance of my Research Colleague 1st Lt. Jess Scott. Most of the work reported here is the result of a joint effort with him and I greatly appreciate his help with this project. I would also like to thank Dr. Phil Yu for assistance during the early stages of this project.

I. PROLOGUE

In the past several years, the need for developing better cathodes has been realized. The expected demand for cathodes that have the extreme lifetimes required for satellite applications, the high current densities necessary for radar and millimeter wave applications, and improved reliability for all applications has led to increased research in numerous types of cathodes. The Avionics Laboratory has particular interest in nickel matrix cathodes especially the Medicus cathode. It is hoped that these cathodes can fulfill the emission requirements of microwave tubes in the near future. This report describes a facility under development for testing cathodes and investigating some of their important properties.

II. OVERALL OBJECTIVES

After preliminary experimentation and familiarization with the problems associated with cathode characterization, two related objectives were defined. We decided to concentrate our efforts on improving, refining and in part automating the thermionic emission facilities that were available in this laboratory. These facilities are to be used in the normal characterization of cathodes. In addition they are to be used in conjunction with experiments to learn more about the underlying physics of the activation and emission processes that occur in the cathodes. Along these lines it was decided to design an experiment to study the poisoning effects of various gases on the cathodes. In summary, the objectives are

1. to provide an improved facility for thermionic emission studies
2. to examining the effects of various gases on the performance of the cathodes

III. BACKGROUND

A. The Medicus Cathode

This cathode is a high work function metal with an interdispersed low work function emissive material. It is designed to furnish high values of continuous current at relatively low temperatures. By virtue of its metal - nickel - matrix it is rugged and can be easily fabricated and shaped. Its fabrication is characterized by a rolling and annealing cycle. Preimpregnated nickel and alkaline earth carbonates are sintered on a nickel base and rolled to about half the thickness in several steps. Between each step the cathode is annealed in hydrogen.

To complete the processing of the cathode, it must be activated. This is usually accomplished in two steps. The cathode is heated in vacuo until the carbonates are reduced to oxides. This is usually completed by the time the system has reached 800°C to 850°C . The second step consists of holding the cathode at approximately 850°C and drawing current by applying a voltage across the cathode and an anode arranged together as a diode. Sufficient voltage is applied so that the diode is operating in the transition region between the temperature limited and space charge limited operating regimes. For a good cathode, the current will increase with time eventually reaching a fixed value. The voltage is then increased so that the diode is again operating in the transition region. When the current has readjusted to a new, constant value, the voltage is again increased.

This process is repeated until no further increase in current occurs. The cathode is activated.

B. Thermionic Emission

Thermionic emission is the process whereby electrons escape from a hot metallic surface into a vacuum. If the surface is used as a cathode and all the emitted electrons are collected by an anode, the cathode is said to give saturated emission. The current density in this case is called the saturated current density J_s . The electrons that escape from a metal are those whose initial energy was about equal to the Fermi energy. If these electron acquire an additional amount of energy ϕ , called the work function, they can just escape the metal. The relation between the saturated current density J_s and this "barrier height" ϕ is given by the Richardson equation

$$\begin{aligned} J_{so} &= \frac{4\pi emk^2}{h^3} T^2 \exp\left(-\frac{\phi}{kT}\right) \\ &= A T^2 \exp\left(-\frac{\phi}{kT}\right) \end{aligned}$$

where the symbols have their standard definitions. Quantum mechanically, a small fraction of the electrons incident on the surface of the metal from within will be reflected so that the equation above should be multiplied by a transmission coefficient. The theoretical value of A is rarely obtained in actual practice partly because the work function ϕ is temperature dependant. A common practice is to insert the theoretical value of A into the equation and to regard the equation as the definition of the effective work function of the material ϕ^1 .

Once an electron escapes from a surface, it is strongly attracted back to that surface by the image positive charge. To collect the electrons then, an electric field must be imposed. For low values of the field, only a fraction of the electrons are collected, the remainder forming a space charge

cloud near the cathode. One talks of space charge limited emission in this case. The current density calculation must include the contribution of this space charge to the potential. This requires the solution of Poisson's equation for the particular geometry used. An approximate solution for a planar diode is

$$J = \frac{4}{9} \epsilon_0 \left(\frac{2e}{m} \right)^{1/2} \frac{V^{3/2}}{d^2} .$$

This is Child's Law. V is the anode voltage and d the cathode - anode spacing. A plot of $J^{2/3}$ vs V is often used to verify diode performance.

The presence of an external field counteracts the image charge force on escaping electrons. This has the apparent effect of reducing the work function of the material. This is the Schottky effect whereby the saturated emission current is a function not only of the temperature and work function but of the applied field. The emitted current in the saturated region is given now by

$$J_s = J_{so} \exp \left[\frac{eq\sqrt{E}}{kT} \right]$$

where E is the field strength and q is a geometrical factor. This indicates that the Richardson equation holds only for zero applied field. To obtain effective work functions the above equation must be used to obtain the zero field current densities which are then used in the Richardson equation.

A number of complications have been ignored in the above discussion. Patch effects, work function differences for different crystal planes, semi-conducting surfaces and surface states are additional concerns which would lead us far astray here. Suffice it to say that the above equations are useful in organizing and comparing emission data for our cathodes.

IV. INITIAL EXPERIMENTS

Some preliminary experiments were performed on cathodes using the optical and electrical measurement facilities that were available.

From a four point measurement of a piece of cathode material a resistivity value of $2.3 \times 10^{-5} \Omega\text{-cm}$ was deduced. This is decidedly metallic! Current emission theories emphasize the semiconducting nature of the emitting surface. Apparently this surface is quite thin and/or recessed with metallic nickel peaks making contact with the measuring circuit thus shorting out the semiconductor portions of the surface. An optical experiment was then tried. Eight samples were subjected to ultra violet light while at 77°K . No luminescence was observed. Three of these samples were further studied at 4°K under illumination by a He-Cd laser which radiates at 3150 \AA ($\sim 3.8 \text{ ev}$). Again no luminescence was observed between 4000 \AA to 1.2 microns. This is presumably due to the irregular nature of the surface and internal structure and a high probability for non-radiative transitions. These lines of inquiry did not appear to be too promising and so these studies were discontinued in favor of the thermionic emission studies.

A large number of parameters are involved in the cathode activation and emission processes. We have made an effort to improve our knowledge and control of some of these parameters. One of the problems encountered in thermionic emission studies is accurate determination of the cathode temperature which must be well known for theoretical analysis. The cathode is generally not visible when in operation and the temperature is often inferred from optical pyrometer measurements of the nickel sleeve upon which the cathode is mounted. Corrections must be made for the emissivity of the nickel and absorption of the glass envelope. Using optical pyrometer measurements of cathodes and sleeves in combination with thermocouple measurements we find that the glass wall reduces the apparent temperature by about four degrees C while the nickel

sleeve temperatures are about 40°C less than the cathode temperature. We can now make reliable, consistent temperature measurements.

V. MASS SPECTROMETER

In order to obtain more detailed information about the processes occurring during the activation and operation of a cathode, we have incorporated a mass analyser into the system. An older G.E. model monopole partial pressure analyses was available. Its performance had seriously degraded. However, we were able to use the electron gun and monopole analyser portions. To these we added a Kiethley 417 High Speed Pico-ammeter to amplify the current from the analyser, a Wavetek 146 Multifunction Generator to sweep the analyser voltage, a Houston Instruments chart recorder and a Tektronics 564 Storage Oscilloscope for recording of the spectra. At high pressures (10^{-7} torr range) where there are strong signals, we can record a mass scan from 0 to 60 Amu in about 90 seconds with good response. Such speed is useful when following the evolution of various gases during the activation process. Slower scans with some filtering are possible when signal to noise improvement is required. Two examples are given in Figures 1 and 2. The mass number, starting from zero, increases to the right along the abscissa while the signal amplitude in arbitrary units is displayed along the ordinate. The first trace shows the residual gases in a system that had been pumped for a long time. The second is for the same system some time after the cathode in the sample chamber had been activated. The same gases are present but their proportions are different. Interpretation of these patterns require that calibrations be done with the various gases to obtain their cracking patterns under the existing conditions. Some calibrations have been done using the vacuum system to be described later.

AMPLITUDE (ARBITRARY UNITS)

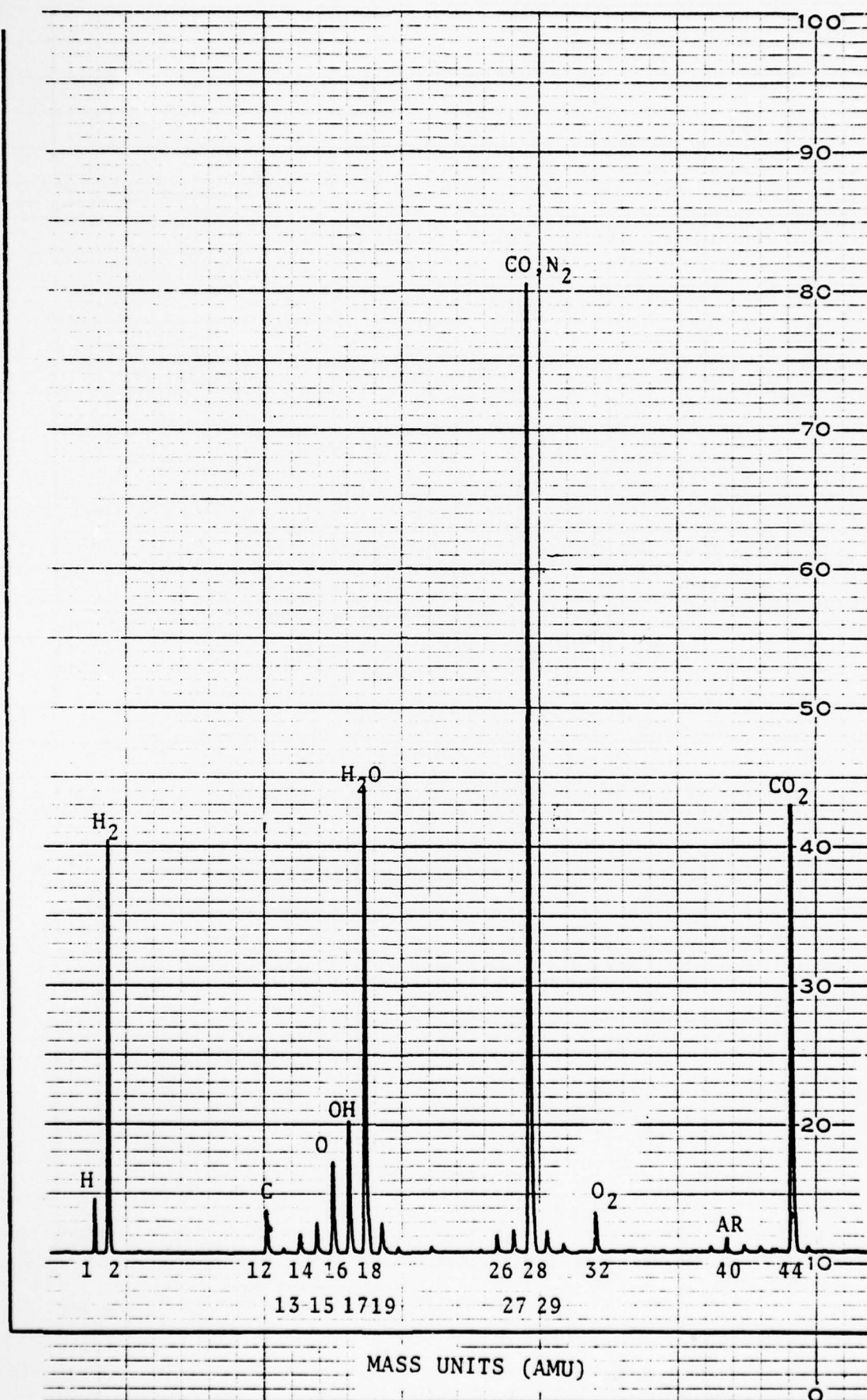


Figure 1. Mass spectra of residual gases before activation

AMPLITUDE (ARBITRARY UNITS)

Figure 2. Mass spectra of residual gases after activation

CO, N_2

CO_2

H

H_2

OH

O

H_2O

MASS UNITS (AMU)

44

32

28

18

17

16

15

14

13

12

11

10

9

8

7

6

5

4

3

2

1

We have studied one cathode extensively using the mass analyser. A schematic diagram of the system is shown in Figure 3. A cathode was activated over a period of a week and its emission characteristic studied. During activation, the cathode temperature, total gas pressure and gas composition could be monitored. Very noticeable changes in the gas composition were observed during the course of the activation process. In particular, mass peak 44 increased two orders of magnitude then decreased back to its pre-activation (residual) level. This seems to be reasonable since carbonates are being broken down to oxides with the evolution of gases CO or CO₂. However published mass spectra of CO₂ indicates that it is cracked with a CO peak (mass 28) being the predominant peak not a mass 44 peak. This is puzzling. In addition, after the mass 44 peak subsided, a quite strong mass 28 peak dominated the spectrum. This is not the case for a residual gas spectrum. This behavior can be noticed by comparing Figure 1 with the oscilloscope photographs in Figure 4. The upper photograph was taken during activation. The lower trace shows the strong peak at mass 44. (The upper trace is a 5 x enlargement of the lower trace) The pressure in the diode and temperature increases, the CO₂ peak increases and then decreases until at 930°C the composition of the gases arrears as in the upper trace of the lower figure. The CO peak dominates. The lower traces in the photo are for lower temperatures and pressures as the cathode temperature is reduced.

With the recent assembly of the revamped vacuum system shown in the next section, some calibration data has been obtained for CO, CO₂ and O₂. These are shown in Figures 5, 6, and 7.

The CO spectra shows the mass 28 - CO peak superposed on the residual gas spectra. The O₂ spectra show a strong O₂ peak at 32 Amu and a relatively strong O peak at 16 Amu. In addition a peak at 28 occurs which must

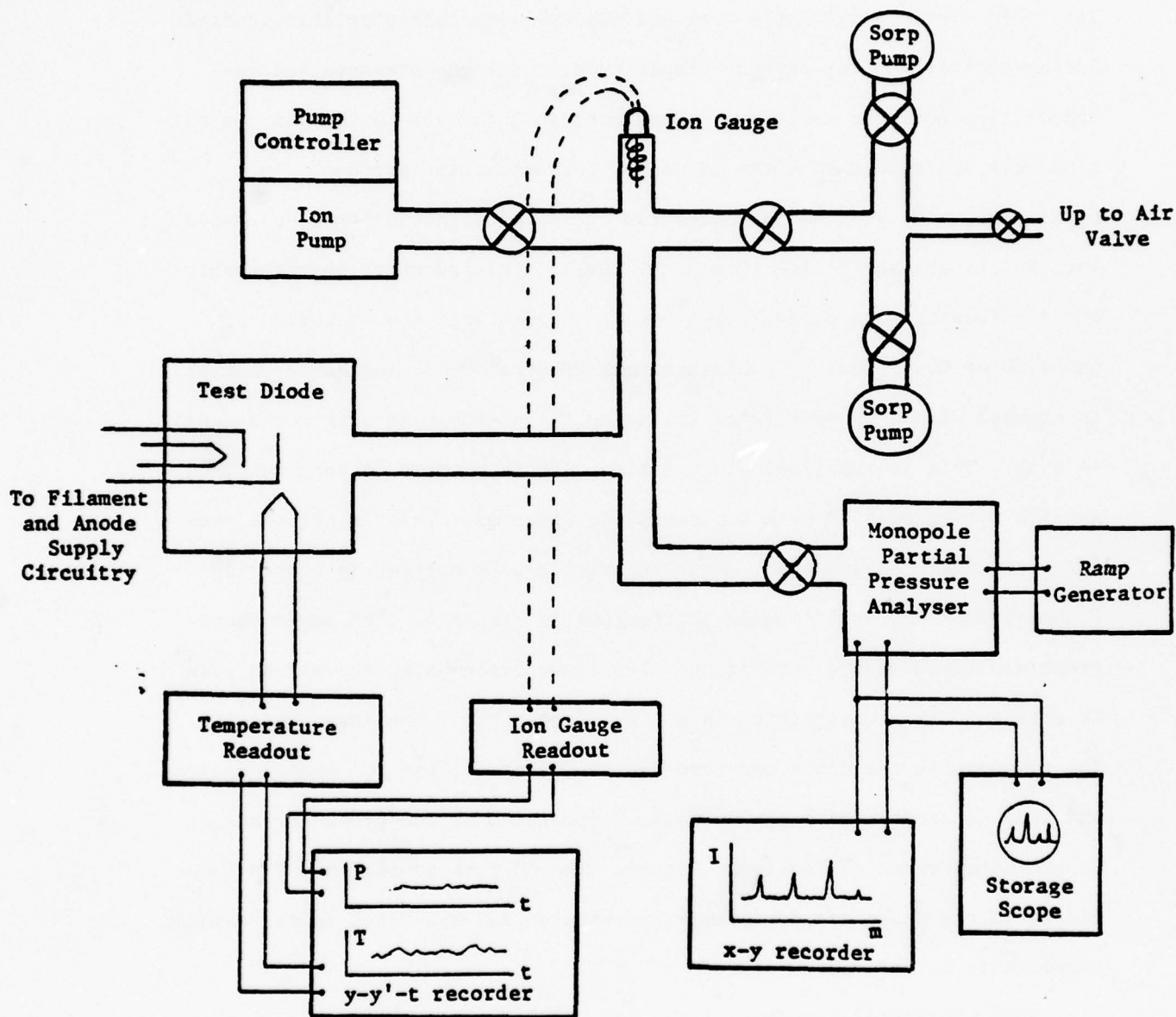


Figure 3. SCHEMATIC OF SYSTEM USED FOR CATHODE ACTIVATION

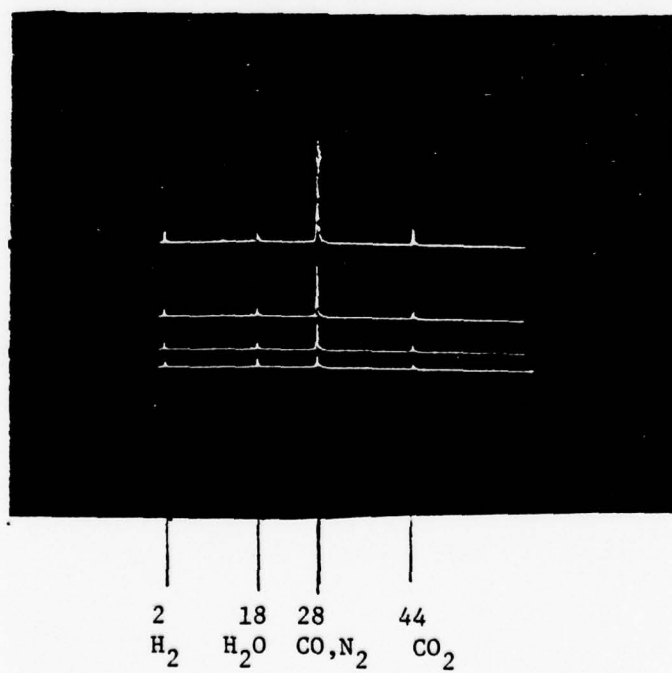
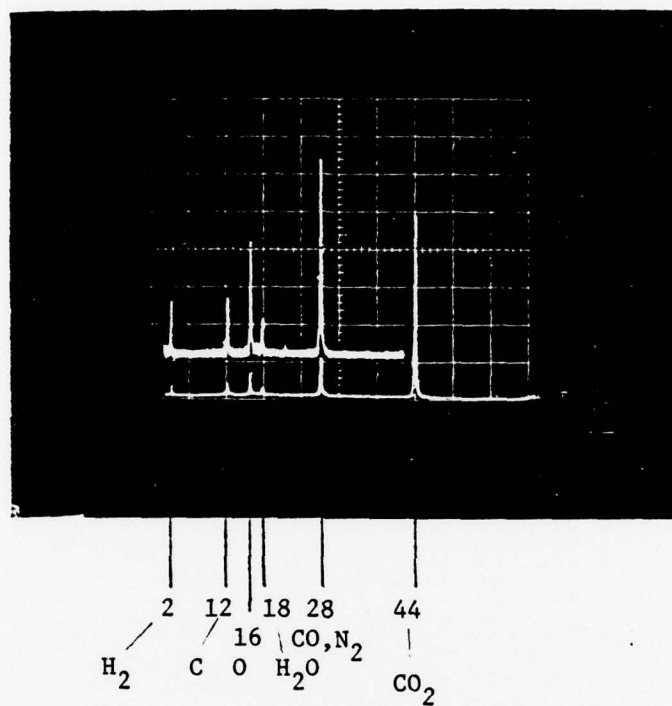


Figure 4. Mass spectra of gases during activation

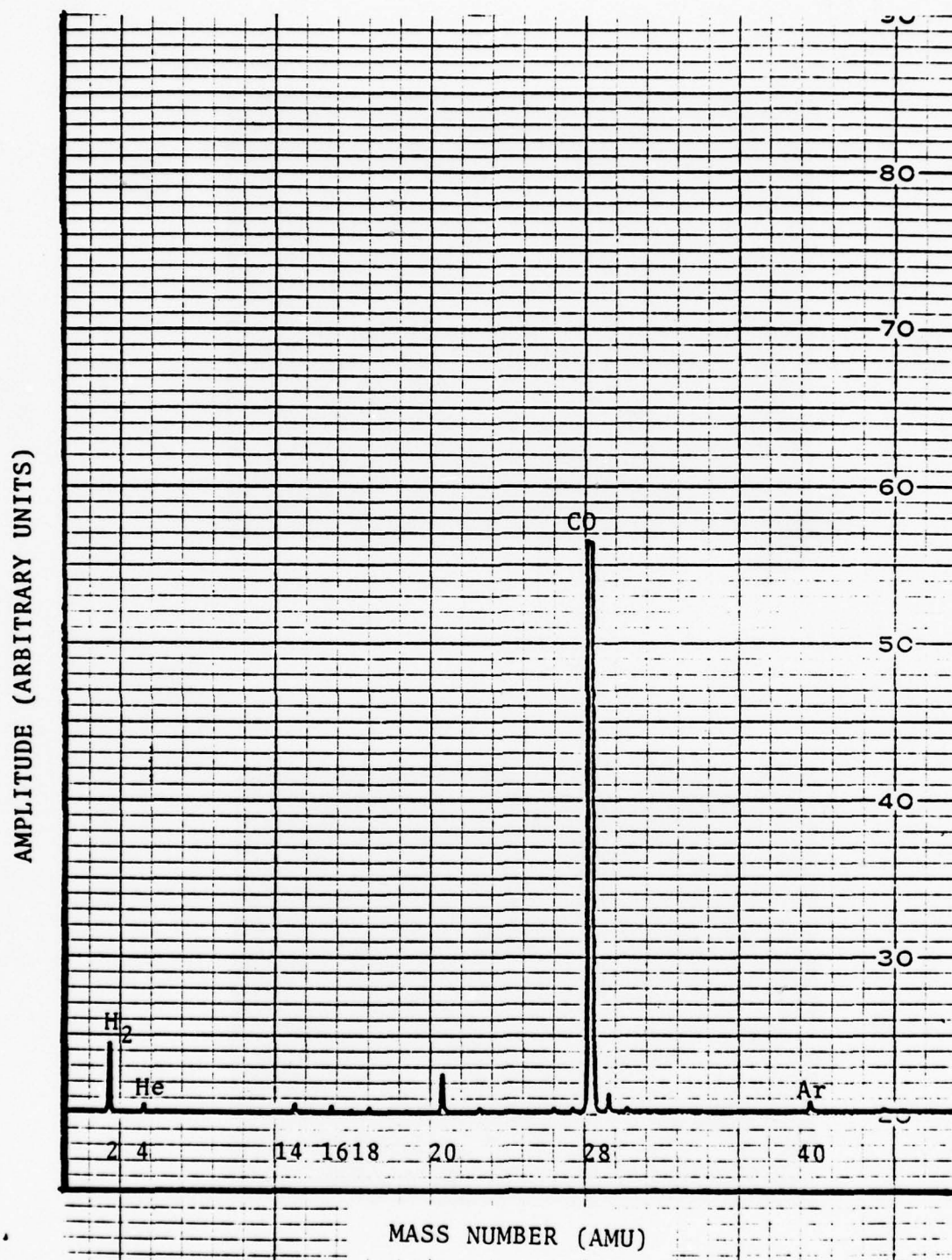


Figure 5. Mass spectra of carbon monoxide (CO)

AMPLITUDE (ARBITRARY UNITS)

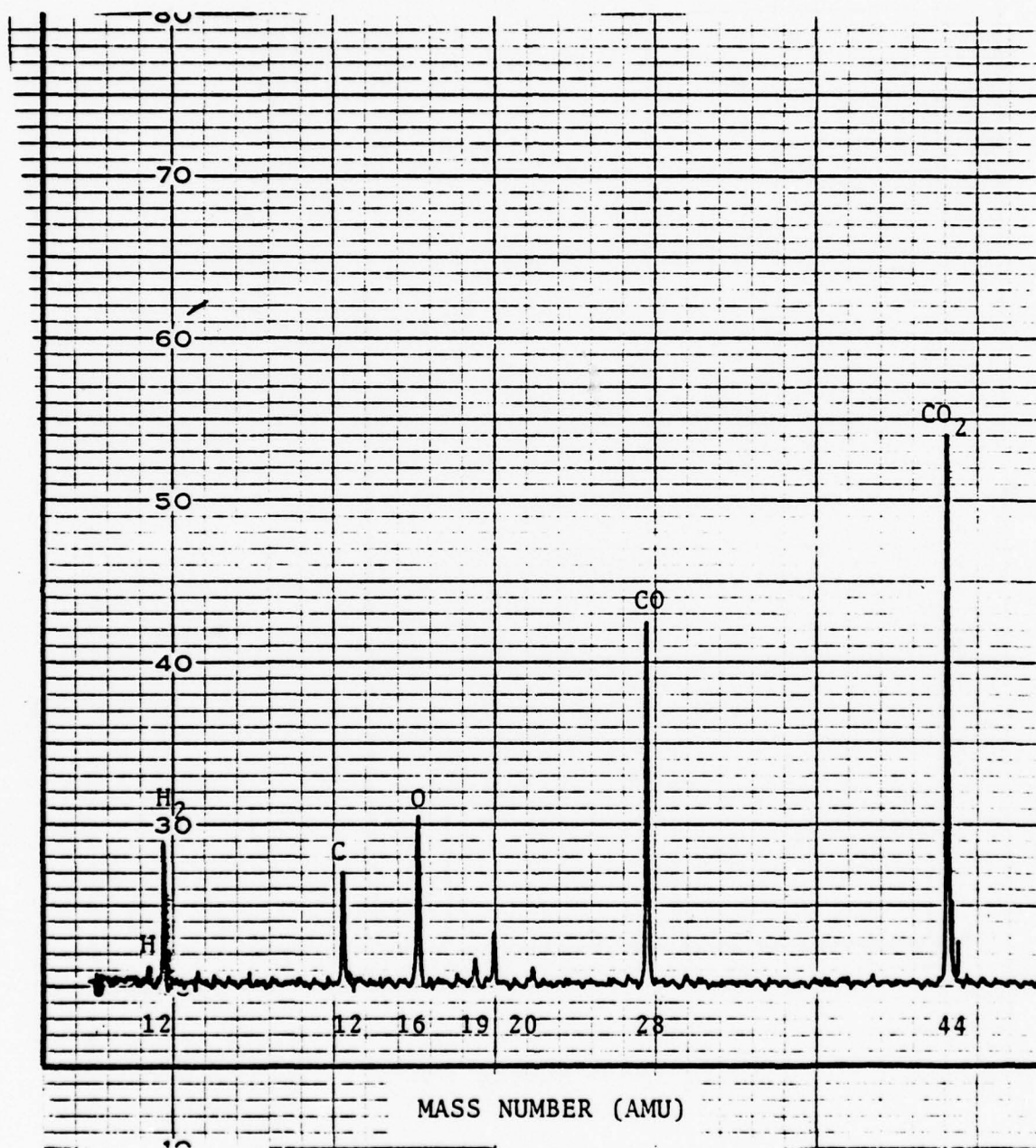


Figure 6. Mass spectra of carbon dioxide (CO₂)

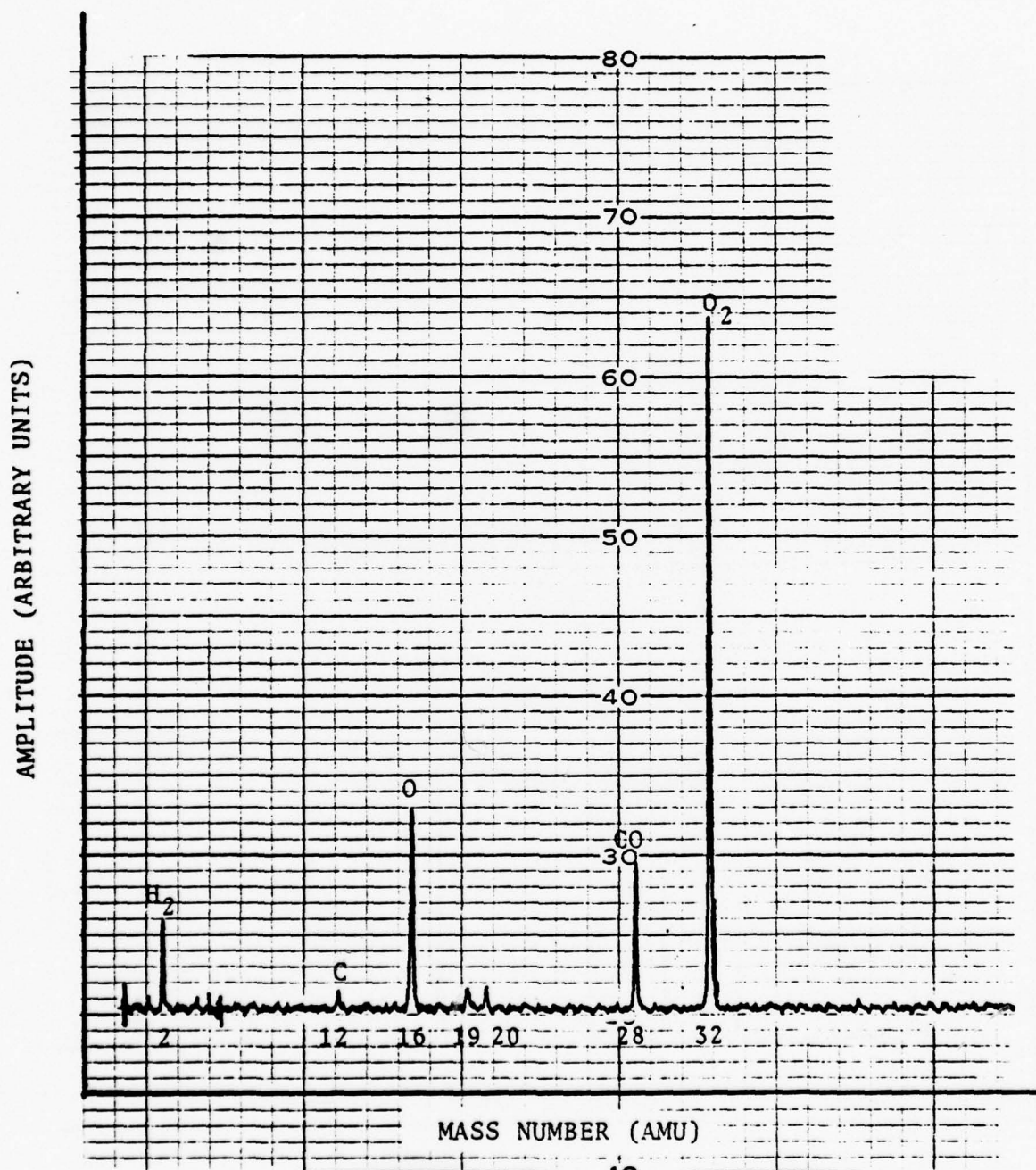


Figure 7. Mass spectra for oxygen (O_2)

be in part due to N_2 but also is probably due to some residual CO in the vacuum system. (The CO spectra were taken before the O_2 spectra). The CO_2 spectra exhibits CO_2 , CO, C and O peaks.

These calibration spectra seem to substantiate the statement above that after the evolution of CO_2 during activation, here is a great deal of CO present.

The emission properties of the cathode were not particularly good and it seemed quite susceptible to poisoning. Some authors have reported reactivation of cathodes when operated in an argon discharge. This cathode did not respond to such treatment.

A number of experimental problems became evident during this experiment. The activation process took almost a full week with constant monitoring and recording of data. Overnight operation and more efficient data gathering is needed. These and other considerations led to the restructuring of the system as described in the next section.

V. MEASUREMENT SYSTEM FOR THERMIONIC EMISSION STUDIES

A. Instrumentation

Initially, the current-voltage characteristics of a diode containing a test cathode were obtained using the circuit configuration in Figure 8. None of the data taking was automated. Data could be read from the meters and/or the resulting graph produced by the recorder or oscilloscope in the case of pulsed operation of the high voltage supply.

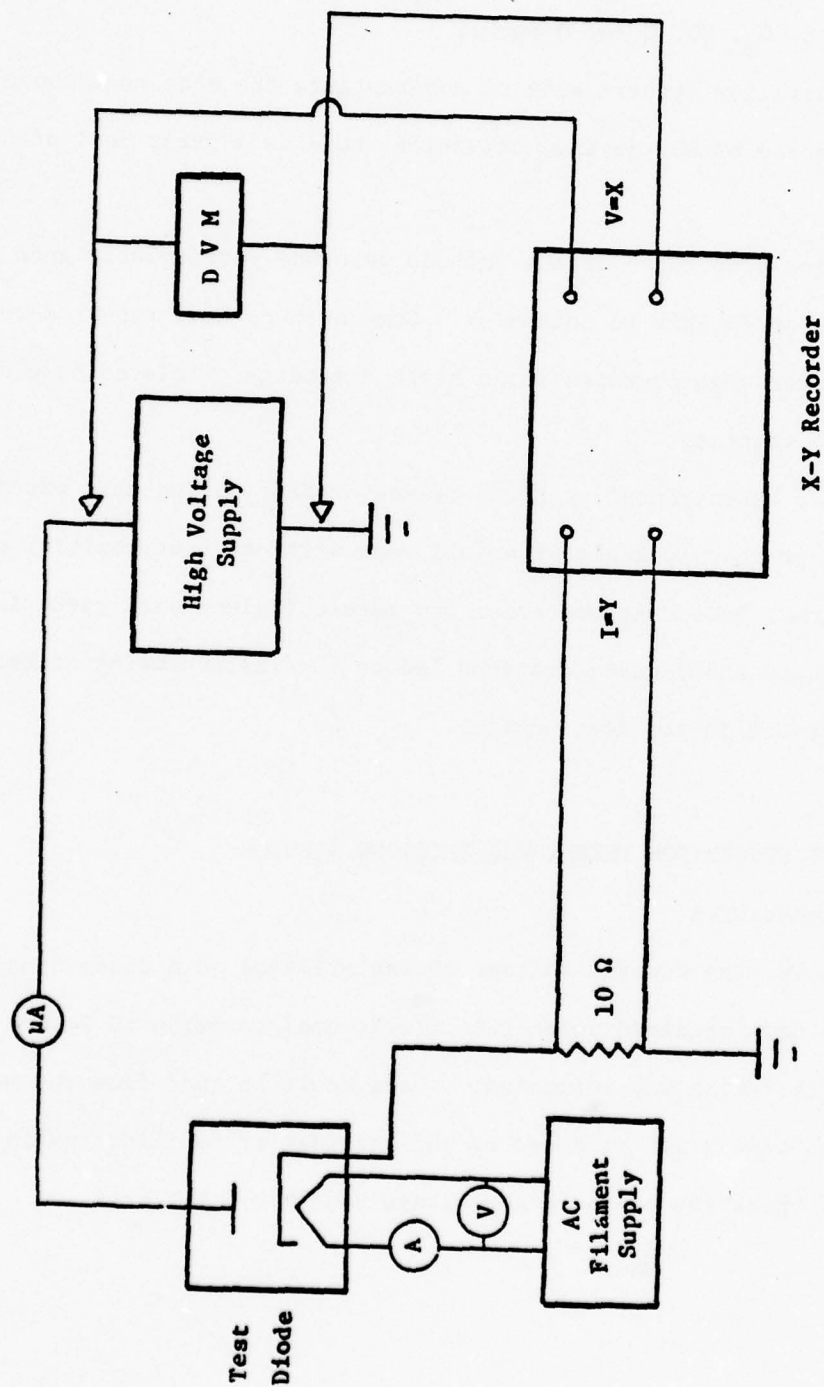


Figure 8. CIRCUIT DIAGRAM FOR OBTAINING I-V CHARACTERISTICS OF DIODES

The next figure, Figure 9, shows the system as presently configured. The filament supply can either be a constant voltage/constant current DC supply or a constant temperature supply. This latter supply is a redesigned version of a supply constructed in this laboratory and for reasonably large temperature perturbations can maintain the temperature within $\pm 1^{\circ}\text{C}$. This is important since isothermal current-voltage curves can best be handled theoretically.

The test diode will usually be one with a water cooled anode capable of holding eight cathodes for test. Diodes for specialized experiments can be used or the eight cathode diode can be reconfigured for example, to hold fewer cathodes and thermocouples for direct temperature measurement.

The voltages and currents that were read from meters or graphs in the previous case can now be fed into a data logging system and can be presented as numerical lists and/or as a graphical display. We can keep track of the many variables involved and with the calculator can graph them in various ways. At present we are using the system on line. The data is received, plotted and not retained. An example of the output of the plotter is shown in Figure 10. The cathode in use had gone through the first activation step. The figure records the changes in the parameters during the current activation step for a period of about seven hours. Other combinations of the parameter can of course be presented depending upon the program entered into the calculator.

There are a number of limitations to the system at present. More memory would be desirable. Control functions are lacking. High voltage pulse data can not be handled. These latter two are important limitation. A relay boards must be obtained or constructed that would allow control of the independent

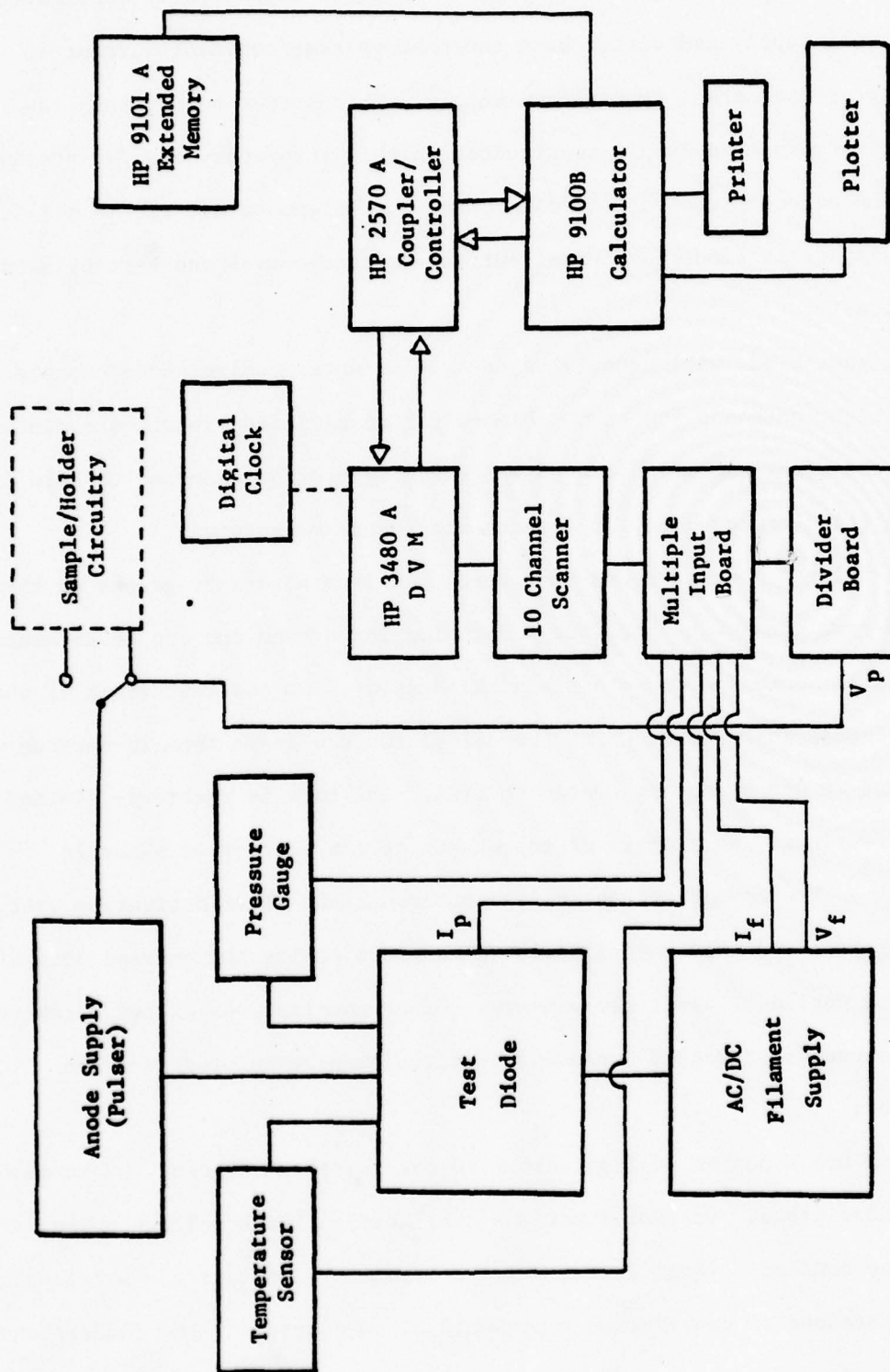


Figure 9. BLOCK DIAGRAM OF MEASURING SYSTEM

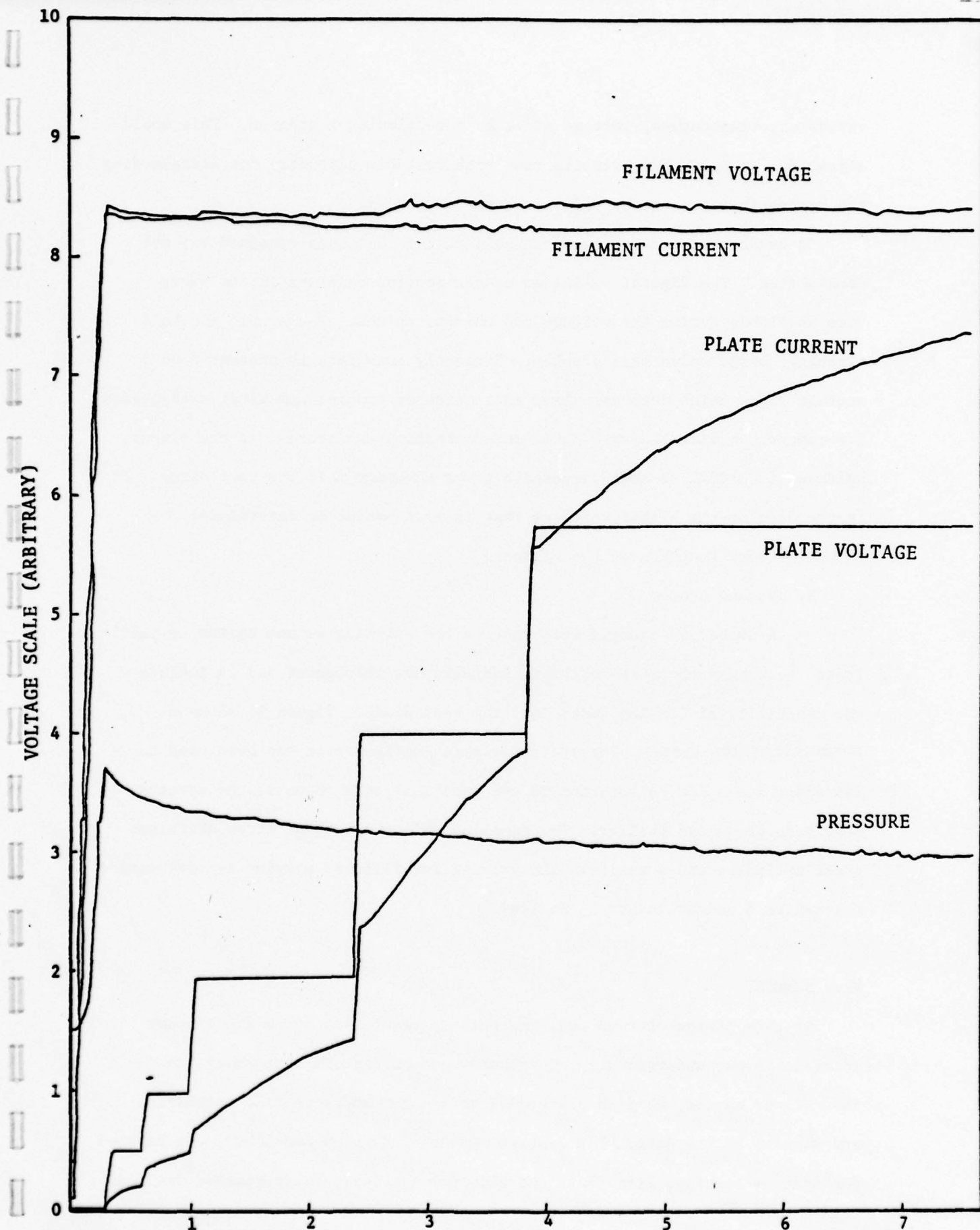


Figure 10. Record of current activation of cathode WB-16

variables, temperature, voltage etc., by the calculator program. This would allow, for example, for overnite runs with suitable circuitry for safeguarding the system in the event of unforeseen circumstances.

To handle pulse data, a sample/hold circuit has been designed but not constructed. The digital voltmeter cannot acquire readings in the short time available during the voltage and current pulses. The sample and hold circuitry would solve this problem. Presently such data is presented on a storage scope which does not allow extraction of precise numerical data needed for comparison with theory. Pulse measurements are performed at the higher voltages ($\sim 1900\text{V}$) to avoid excessive power dissipation in the test diode. It is the data at the higher voltages that is most useful in determining the effective work function of the cathode.

B. Vacuum System

A number of changes were made to the existing vacuum system to facilitate the use of the mass analyser, increase the throughput and to include the capability of leaking gases into the test diode. Figure 11 shows a schematic of the system. The system in this configuration has been used to introduce gases for calibration of the mass analyser. Some of the spectra have been presented earlier. The sampling unit consists of three stainless steel cylinders and a manifold and valving for filling, purging and evacuating the manifold and cylinders as desired.

VI. SUMMARY

As this project turned out, the primary result was not a significant increase in our understanding of cathodes but an instrumentation system that allows easier, more complete data collection and versatile processing and display of the data. The system, with the addition mentioned, can be used for routine testing, life tests and also for specialized experiments designed

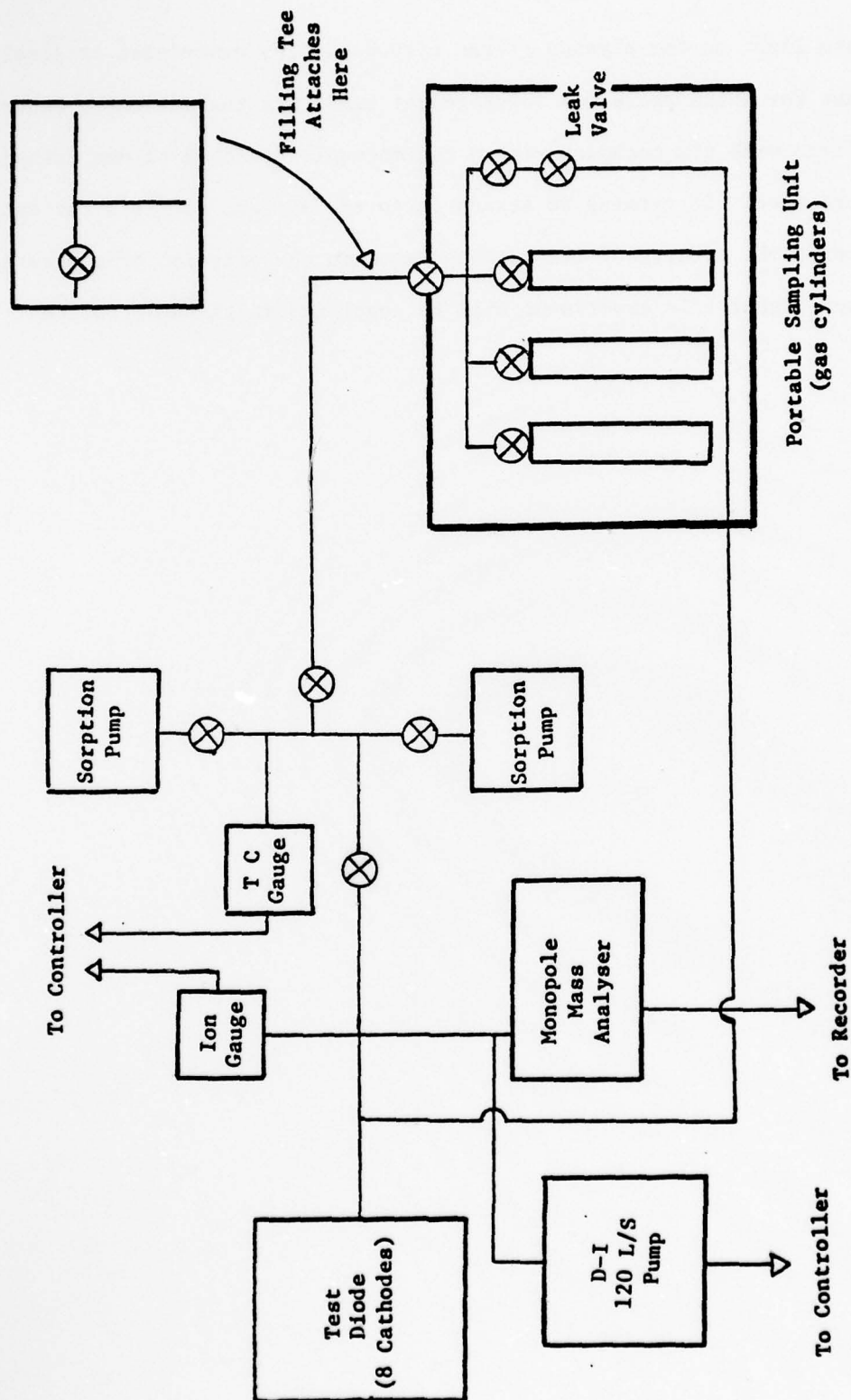


Figure 11. DIAGRAM OF VACUUM SYSTEM (⊗ DENOTES VALVE)

to shed more light on the physics of the cathodes. The experiment originally planned, but for which there was insufficient time, was the poisoning experiment. A test vehicle with six cathodes with a thermocouple attached to one cathode has been prepared. It remains to attach it to the system, activate the cathodes and to observe the effects of the various gases on the emission of the cathodes. It is expected that this experiment will be completed in the near future.

GENERAL REFERENCES

1. Solid State Physical Electronics. A. van der Ziel, Prentice Hall, Englewood Cliffs, N.J. 1976.
2. "Thermionic Emission and Work Function." G.A. Haas and R.E. Thomas in Techniques in Metals Research Vol VI, Part 1. E. Passaglia, Ed. Interscience Publishers, New York 1972.

EFFICIENT FAULT ANALYSIS IN ANALOG CIRCUITS

Alfred T. Johnson, Jr.
Center of Engineering
Widener College
Chester, PA 19013

August 1978

FINAL REPORT: USAF-ASEE SUMMER FACULTY RESEARCH PROGRAM, WPAFB

AIR FORCE AVIONICS LABORATORY
AIR FORCE WRIGHT AERONAUTICAL LABORATORIES
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

EFFICIENT FAULT ANALYSIS IN ANALOG CIRCUITS

Alfred T. Johnson, Jr.
Center of Engineering
Widener College
Chester, PA 19013

Abstract

Fault analysis in analog networks is a form of network parameter identification. The problem of finding network parameters from measurements at the accessible terminals can be expressed as the solution of a system of non-linear equations. Such a system of equations is invariably solved by a multidimensional search. Every step of the search requires solving for the network responses in terms of the parameters, comparing the solution to the measured responses, and then adjusting the parameters in such a way as to produce a response closer to the measured response. If the network is analyzed by nodal equations, the computation of the responses requires inverting a matrix of order equal to the number of inaccessible nodes. The time for this analysis is excessive in practical applications, and poses a major impediment to fault analysis in analog circuits. Using nodal equations for the analysis of open circuits poses no special problems, since an open circuit can be represented by a zero admittance. Short circuits, however, are represented by an infinite admittance, which presents further difficulties in searching for a solution when nodal equations are used.

We have applied a formula of Householder to compute the response of an electrical circuit with either a single open or short circuit. In the short circuit case the matrix inversion is computed with about 1/15 of the normal number of multiplications. In writing the software, sparse matrix techniques were used. This reduced both the storage requirements and the execution time. Examples are given.

TABLE OF CONTENTS

SECTION	PAGE
I INTRODUCTION	1
II THE USE OF NODAL EQUATIONS IN CATASTROPHIC FAULT ANALYSIS	2
1. Formulating the y-parameters	2
2. Fault Analysis	5
3. Short Circuits	7
3.1 Forming the y-Parameters of the Shorted Network	9
3.2 Finding the Inverse of Y'_2 Efficiently	10
4. Open Circuits and Other Large Parameter Changes	13
5. Sparse Matrix Techniques	14
5.1 Evaluating the Inverse of Y'_2	14
5.2 Evaluating the y-Parameters	18
5.3 Summary of Efficient Computation	22
6. Status of Current Software	23
6.1 Using the Software	24
6.2 Example of Software Use	28
6.3 Effectiveness of Software in Locating Short Circuit Faults	32
III CONCLUSIONS	34
IV RECOMMENDATIONS	35
APPENDIX A COMPUTER PROGRAM LISTINGS	36
REFERENCES	54

I INTRODUCTION

It is perhaps surprising that although the problem of fault isolation in analog circuits has been studied longer than the comparable problem in digital circuits¹, the digital problem is much better understood, and automatic test equipment (ATE) for isolating faults in digital equipment to the chip and even to the gate level is available, but ATE available for isolating faults in analog circuits is much less sophisticated[15-17]. The explanation of this phenomenon lies partly with the fact that analog signals are inherently more complex than digital signals. They occur continuously in time, rather than at discrete times, and their values (in principle) have infinite resolution, instead of being truncated to a finite number of bits.

In the linear analog circuits discussed in this report, the problem of signal variety can be partly solved by restricting observation to the sinusoidal steady state, since the system behavior to any input can be determined from a knowledge of the sinusoidal steady state response at all frequencies. Clearly all frequencies cannot be considered, and the problem of how to choose a set of frequencies has only begun to receive attention [13,14]. In addition, the response at any frequency is related to the parameter values and the network graph by non-linear equations, to which the only method of solution possible is a time consuming search.

In this report we show how the search for single short circuits can be accomplished efficiently enough to make the computation practical. We also demonstrate the effectiveness of this method in identifying short circuit faults in analog circuits whose resistance, capacitance, inductance, and gain values differ significantly from their nominal values.

¹ An extensive bibliography of both digital and analog fault isolation is given by R. Saeks and S.R. Liberty [12].

II THE USE OF NODAL EQUATIONS IN CATASTROPHIC FAULT ANALYSIS

1. Formulating the y-Parameters

We will consider linear, lumped-parameter networks having $p+1$ accessible terminals, numbered 0 through p , as shown in Fig. 1. The

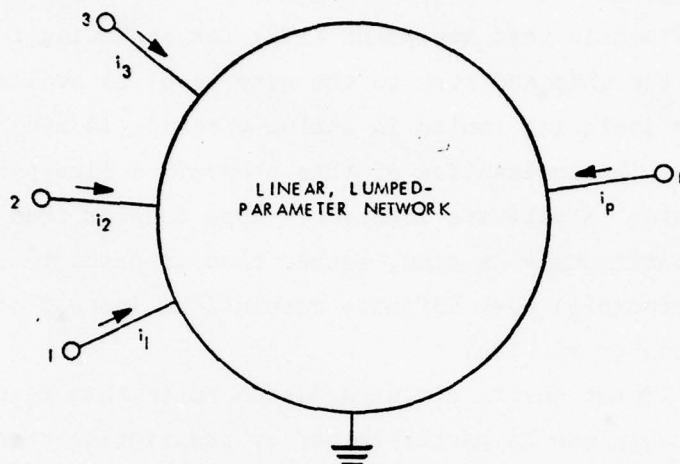


Fig. 1 Linear, lumped parameter network with $p+1$ accessible terminals

network of Fig. 1 contains $n+1$ nodes, including the reference node, and b branches and is assumed to contain no internal independent sources but may contain dependent sources. We will formulate the nodal equations using the standard branch shown in Fig. 2.

The passive element in branch k , denoted by its admittance Y_{bk} , must be non-zero, but any of the sources may be zero or all may be present, if desired. In the equations that follow, \underline{v} is the branch voltage vector, \underline{i} is the branch current vector, \underline{e} is the node voltage vector, \underline{v}_s is the independent voltage source vector, and \underline{i}_s is the independent current source vector, using the notation of Desoer and Kuh [1]. The dimension of each of these vectors is $b \times 1$. Considering the k th standard branch shown in Fig. 2, it is clear that the branch $v-i$ constraints are given by

$$\underline{i} = \underline{i}_s + (\underline{G}_m + \underline{Y} - \underline{Y}_m)\underline{v} + (\underline{\alpha} - \underline{Y}_m \underline{R}_m) \underline{i} - \underline{Y}_m \underline{v}_s \quad (1)$$

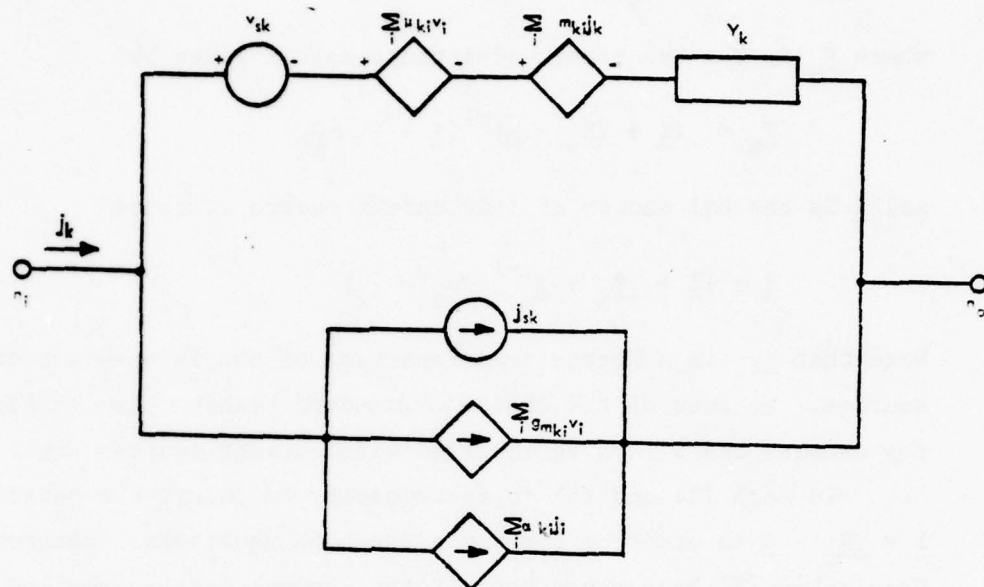


Fig 2 Standard branch k

where $\underline{Y} = \text{diag}(Y_k)$ is the branch passive admittance matrix, and \underline{G}_m , $\underline{\mu}$, $\underline{\alpha}$, and \underline{R}_m are the coupling matrices due to the controlled sources. The dimension of each of these matrices is $b \times b$. (The elements of matrices \underline{G}_m and $\underline{\alpha}$ are current sources, and the elements of matrices $\underline{\mu}$ and \underline{R}_m are voltage sources controlled respectively by voltages and currents). Following the convention of Desoer and Kuh, define the node-to-branch incidence matrix \underline{A} ,

$$\{\underline{A}\}_{ij} = \begin{cases} +1, & \text{if branch } j \text{ leaves node } i \\ -1, & \text{if branch } j \text{ enters node } i \\ 0, & \text{if branch } j \text{ is not incident on node } i \end{cases}$$

Solving (1) for \underline{j} and using the Kirchhoff current law $\underline{A}\underline{j} = 0$ and the Kirchhoff voltage law $\underline{A}^T \underline{e} = \underline{v}$, it can be shown that the network node

equations are

$$\underline{A} \underline{Y}_b \underline{A}^T \underline{e} = \underline{i} \quad (2)$$

where \underline{Y}_b is the bxb branch admittance matrix given by

$$\underline{Y}_b = (\underline{1} + \underline{YR}_m - \underline{\alpha})^{-1} (\underline{Y} - \underline{Y}_\mu + \underline{G}_m) \quad (3)$$

and \underline{i} is the nx1 vector of independent source currents

$$\underline{i} = (\underline{1} + \underline{YR}_m - \underline{\alpha})^{-1} (\underline{Yv}_s - \underline{j}_s) \quad (4)$$

Note that \underline{Yv}_s is a Norton transformation of the independent voltage sources. Because of the choice of standard branch shown in Fig. 2, the network can always be modelled with current sources only.

In both (3) and (4) it is necessary to invert the matrix $\underline{1} + \underline{YR}_m - \underline{\alpha}$ in order to formulate the node equations. Johnson and Pennington [2] have shown that if the network can be modelled so that the branch current j_k of a branch which contains a current controlled source does not control any source, then the inverse of $\underline{1} + \underline{YR}_m - \underline{\alpha}$ is $\underline{1} - \underline{YR}_m + \underline{\alpha}$. Since most networks can be modelled this way, forming the node equations does not present any difficulty.

According to the convention chosen, nodes 0 through p are accessible, and nodes p+1 through n are inaccessible. Partition the rows of the node-to-branch incidence matrix \underline{A} into accessible and inaccessible nodes

$$\underline{A} = \begin{bmatrix} \underline{A}_1 \\ \text{---} \\ \underline{A}_2 \end{bmatrix} \quad (5)$$

where \underline{A}_1 is the pxn incidence matrix of accessible nodes, and \underline{A}_2 is the qxn incidence matrix of inaccessible nodes, and $q = n-p$ is the number in inaccessible nodes. Using (5) in (2) and partitioning \underline{i} and \underline{e} as follows

$$\underline{i} = \begin{bmatrix} \underline{i}_1 \\ \hline \underline{i}_2 \end{bmatrix}, \quad \underline{e} = \begin{bmatrix} \underline{e}_1 \\ \hline \underline{e}_2 \end{bmatrix}$$

where \underline{i}_1 and \underline{e}_1 are $p \times 1$ vectors, and \underline{i}_2 and \underline{e}_2 are $q \times 1$ vectors, we obtain

$$\begin{bmatrix} \underline{A}_1 \underline{Y}_b \underline{A}_1^T & \underline{A}_1 \underline{Y}_b \underline{A}_2^T \\ \hline \underline{A}_2 \underline{Y}_b \underline{A}_1^T & \underline{A}_2 \underline{Y}_b \underline{A}_2^T \end{bmatrix} \begin{bmatrix} \underline{e}_1 \\ \hline \underline{e}_2 \end{bmatrix} = \begin{bmatrix} \underline{i}_1 \\ \hline \underline{i}_2 \end{bmatrix} \quad (6)$$

Now eliminate \underline{e}_2 from (6) and make use of the fact that $\underline{i}_2 = \underline{0}$, since no independent sources are connected to inaccessible nodes, to obtain

$$\underline{i}_1 = \underline{A}_1 \underline{Y}_b [\underline{1} - \underline{A}_2^T (\underline{A}_2 \underline{Y}_b \underline{A}_2^T)^{-1} \underline{A}_2 \underline{Y}_b] \underline{A}_1^T \underline{e}_1 \quad (7)$$

so that

$$\underline{Y}_1 = \underline{A}_1 \underline{Y}_b [\underline{1} - \underline{A}_2^T (\underline{A}_2 \underline{Y}_b \underline{A}_2^T)^{-1} \underline{A}_2 \underline{Y}_b] \underline{A}_1^T \quad (8)$$

is the $p \times p$ matrix of measurable y-parameters defined at the ports having a common "ground" node (node 0 in Fig. 1). The subscript 1 indicates that the y-parameters are evaluated at frequency ω_1 .

2. Fault Analysis

Let $\hat{\underline{Y}}_1$ be the matrix of measured y-parameters, where the measurement is made at frequency ω_1 . A fault represented by a shift of the network element values from their nominal values could be identified by solving the non-linear equation

$$\hat{\underline{Y}}_1 = \underline{Y}_1(\underline{Y}_b) \quad (9)$$

for the unknown parameter values. If a single frequency is used, as in (9), the number of network parameters usually exceeds the number

of independent y-parameters, so that (9) has no unique solution. One approach, suggested by Ransom and Saeks [3], is to find the solution to (9) that minimizes a norm $\|\underline{y}_{bo} - \underline{y}_b\|$, where \underline{y}_{bo} is evaluated from the nominal network element values. Ransom and Saeks give an approximate solution to this problem using linear methods. The difficulty with this approach is that the solution assumes, roughly speaking, that the most likely state of the element values in the network is the one which causes the smallest drift from the nominal values consistent with the measurements. This assumption excludes the possibility of catastrophic faults (open and short circuits).

Another approach is to augment (9) by measurements at a sufficient number of frequencies that a unique solution exists. The relationship between fault resolution and the number of frequencies and measurements is discussed by Sen and Saeks [9, 10]. Although their algorithm requires choosing a number of frequencies, it is not clear how this is done. Thus we might define:

$$\begin{aligned}\hat{\underline{y}} &= (\hat{y}_1 \hat{y}_2 \cdots \hat{y}_k)^T \\ \underline{y} &= (y_1 y_2 \cdots y_k)^T\end{aligned}\tag{10}$$

and solve the equation

$$\hat{\underline{y}} = \underline{y}(\underline{y}_b)\tag{11}$$

Nonlinear equations, such as (11), are often solved by defining a norm $c = \|\hat{\underline{y}} - \underline{y}\|$, and then making a multidimensional search for min c. The amount of computation required by such a search could be prohibitive. Chen and Saeks [4] show that if only one parameter is changed at a time, the measurable system function can be evaluated without the need to invert a matrix, which is apparently necessary in (8). Their method, is based on a formula given by Householder [5].

In this paper we show how the Householder formula can be used to efficiently analyze networks containing short circuits by considering the change in the network graph. The measurable effects of open circuits and finite parameter changes can be analyzed by considering changes in \underline{Y}_b , and although this is essentially equivalent to the method described by Chen and Saeks, it is given here in the context of nodal analysis so that all parameter changes can be analyzed with one formulation of the network equations.

3. Short Circuits

If two nodes of a network become shorted, the new node-to-branch incidence matrix \underline{A}' is related to the original incidence matrix in a very simple way. If nodes i and j are shorted, then \underline{A}' is formed by replacing row i by the sums of rows i and j , and then deleting row j . If node i is the reference node, then \underline{A}' is formed by deleting row j . These row transformations can be obtained by premultiplying \underline{A} by a matrix \underline{R} , each element of which is either 0 or 1. Fig. 3 illustrates the row transformation matrix describing two different short circuits (one involving the ground node and the other not involving the ground node) in a network with 5 nodes and 7 branches. The incidence matrices are partitioned by the dashed lines under the assumption that nodes 3 and 4 in the original network are inaccessible. Thus the top half of the \underline{A} matrix is \underline{A}_1 and the bottom half is \underline{A}_2 . A similar partition is shown for $\underline{A}' = (\underline{A}'_1 | \underline{A}'_2)^T$, but in the transformed network some of the formerly inaccessible nodes may have been made accessible (by virtue of their having been shorted to an accessible node).

The \underline{R} matrix is the "after-to-before" node transformation matrix, which has a column for every original node in the network and a row for every node after the fault. (Since node 0 does not enter into \underline{A} , it does not appear in \underline{R} either). Thus \underline{R} may be partitioned by columns

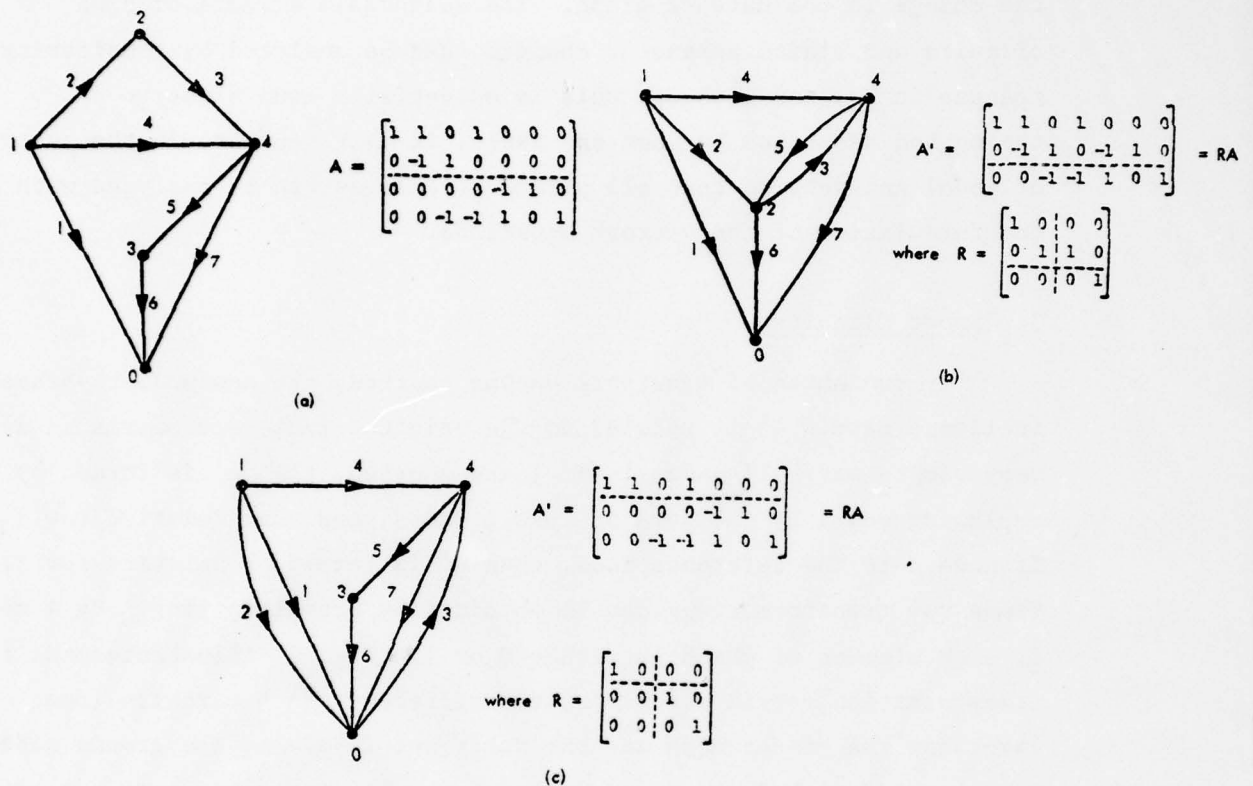


Fig. 3 Effect of short circuits on node-to-branch incidence matrix, A .
 (a) Original network graph. (b) Short circuit between nodes 2 and 3.
 (c) Short circuit between nodes 2 and ground or 0.

and rows into the accessible and inaccessible nodes before (columns) and after (rows) the fault, as shown in Fig. 3.

Denote the partitioning of R by

$$\underline{R} \triangleq \begin{bmatrix} \underline{R}_{11} & \underline{R}_{12} \\ \underline{0} & \underline{R}_{22} \end{bmatrix} \quad (12)$$

where the lower, left partition in (12) is $\underline{0}$ since the faulted network cannot have any inaccessible nodes which were accessible in the unfaulted network. In other words, if a short circuit occurs between an accessible and an inaccessible node, the combined node becomes accessible. Thus

$$\underline{A}' = \underline{R} \underline{A} = \begin{bmatrix} \underline{R}_{11}\underline{A}_1 + \underline{R}_{12}\underline{A}_2 \\ \underline{R}_{22}\underline{A}_2 \end{bmatrix}$$

and therefore

$$\underline{A}'_1 = \underline{R}_{11}\underline{A}_1 + \underline{R}_{12}\underline{A}_2$$

$$\underline{A}'_2 = \underline{R}_{22}\underline{A}_2$$

(13)

3.1 Forming the y-Parameters of the Shorted Network

The y-parameters of networks containing a short circuit can be computed from (8) after replacing \underline{A}_1 by \underline{A}'_1 and \underline{A}_2 by \underline{A}'_2 . Define $\underline{Y}_2 \triangleq \underline{A}_2 \underline{Y}_b \underline{A}_2^T$, which is the qxq nodal admittance matrix of inaccessible nodes. In the analysis that follows, we assume that \underline{Y}_2 is known. (In fact \underline{Y}_2 must be computed only once for any circuit using the nominal element values, so the cost of its computation is a negligible part of the total cost of analysis).

If the short circuit occurs between two accessible nodes, then $\underline{A}'_2 = \underline{A}_2$, and the new y-parameters are found from (8) with no matrix inversion. In fact, if a short circuit exists between two accessible nodes, it can be easily observed by making resistance measurements

between all pairs of accessible nodes, so the method described here is not needed.

If the short circuit involves any inaccessible node, then $\underline{A}'_2 \neq \underline{A}_2$ and the inverse of $\underline{Y}'_2 = \underline{A}'_2 \underline{Y}_2 \underline{A}'_2{}^T$ must be computed. The main contribution of this paper is to show that the inverse of \underline{Y}'_2 may be found in terms of \underline{Y}_2 with about one fifteenth as many multiplications as are required to find the inverse of a general qxq matrix.

3.2 Finding the Inverse of \underline{Y}'_2 Efficiently

From the definitions of \underline{A}'_2 , \underline{R}_{22} , \underline{Y}_2 and \underline{Y}'_2 it is easy to show that

$$\underline{Y}'_2 = \underline{R}_{22} \underline{Y}_2 \underline{R}_{22}{}^T \quad (14)$$

Let us add a row to \underline{R}_{22} as follows

$$\underline{S} \triangleq \begin{bmatrix} \underline{R}_{22} \\ -\underline{u}^T \end{bmatrix} \quad (15)$$

where \underline{u}^T is a row unit vector containing all zeros except for a 1 in the rightmost column representing a shorted node. (At most two columns of \underline{R}_{22} are involved in the short circuit. If only one is involved, then that column of \underline{u}^T is unity. If two columns of \underline{R}_{22} are involved, then only the rightmost one is unity. This convention could be reversed by changing the definition of \underline{R}). We now decompose \underline{S} into \underline{U} and \underline{P}

$$\underline{S} = \underline{U} + \underline{P} \quad (16)$$

where \underline{P} is formed as follows. If the short circuit is between an accessible and an inaccessible node, then $\underline{P} = \underline{0}$. If the short circuit is between two inaccessible nodes, then $\{\underline{P}\}_{i-p, j-p}$ is unity, and the

remaining elements are zero. (In this case, the unit element in \underline{P} is element $\{\underline{R}\}_{ij}$, where nodes i and j are shorted, and $i < j$ due to our numbering convention). With these definitions, it is easy to see that \underline{U} is real and unitary, since its columns are independent unit vectors and thus form an orthonormal set [7]. Consequently $\underline{U}^{-1} = \underline{U}$, and if the short circuit is between an accessible and an inaccessible node, then $\underline{S}^{-1} = \underline{S}$. If the short circuit is between two inaccessible nodes, then \underline{S}^{-1} is only slightly more difficult to compute, as we will show. Now define $\underline{C} \triangleq \underline{S} \underline{Y}_2 \underline{S}^T$, and therefore

$$\underline{C} = \begin{bmatrix} \underline{R}_{22} \\ -\underline{u}^T \end{bmatrix} \underline{Y}_2 \begin{bmatrix} \underline{R}_{22}^T \\ \underline{u} \end{bmatrix} = \begin{bmatrix} \underline{Y}_2' & \underline{R}_{22} \underline{Y}_2 \underline{u} \\ \underline{u}^T \underline{Y}_2 \underline{R}_{22}^T & \underline{u}^T \underline{Y}_2 \underline{u} \end{bmatrix}$$

and

$$\underline{C}^{-1} = (\underline{S}^T)^{-1} \underline{Y}_2^{-1} \underline{S}^{-1} \triangleq \begin{bmatrix} \underline{K} & \underline{L} \\ \underline{M} & \underline{N} \end{bmatrix} \quad (17)$$

Faddeeva [8] shows that $\underline{K} = \underline{Y}_2'^{-1} [1 - (\underline{R}_{22} \underline{Y}_2 \underline{u}) \underline{M}]$, and therefore

$$\underline{Y}_2'^{-1} = \underline{K} [1 - (\underline{R}_{22} \underline{Y}_2 \underline{u}) \underline{M}]^{-1} \quad (18)$$

The inverse on the right side of (18) can be computed efficiently using a formula given by Householder [6,7], which states that if $\underline{A} = \underline{B} - \underline{vw}^T$, where \underline{v} and \underline{w} are column vectors, then

$$\underline{A}^{-1} = \underline{B}^{-1} + \frac{\underline{B}^{-1} \underline{vw}^T \underline{B}^{-1}}{1 - \underline{w}^T \underline{B}^{-1} \underline{v}} \quad (19)$$

where the division on the right side of (19) is by a scalar. Thus the inverse of \underline{Y}'_2 is

$$\underline{Y}'_2{}^{-1} = \underline{K} \left[\underline{1} + \frac{(\underline{R}_{22}\underline{Y}_2\underline{u})\underline{M}}{1 - \underline{M}(\underline{R}_{22}\underline{Y}_2\underline{u})} \right] \quad (20)$$

In order to use (20), \underline{K} and \underline{M} must be found from (17). We have mentioned that if the short circuit is between an accessible node and an inaccessible node, then $\underline{S}^{-1} = \underline{S}^T$. In fact, if the short circuit is between an accessible node and node n , then $\underline{S} = \underline{I}$, and \underline{K} and \underline{M} are simply partitions of (17). If the short circuit is between two inaccessible nodes, we have shown that $\underline{S} = \underline{U} + \underline{P}$, where $\underline{U}^{-1} = \underline{U}^T$, and the rank of \underline{P} is unity. Thus \underline{P} can be written in the form $\underline{v}\underline{w}^T$, and (19) can be used to find \underline{S}^{-1} . Proceeding in this manner it can be shown that $\underline{S}^{-1} = \underline{U}^T + \underline{Q}$, where

$$Q_{rs} = \begin{cases} -1, & \text{if } r = i-p, s = n-p \\ 0, & \text{otherwise} \end{cases}$$

And so, even in the most complex case, \underline{K} and \underline{M} can be evaluated without a matrix inversion. Some time can be saved in computing the triple matrix product in (17) by noting that \underline{L} and \underline{N} (each column vectors) are not needed, so the last column of \underline{C}^{-1} need not be evaluated. Using straightforward matrix multiplication to find \underline{K} and \underline{M} from (17) and then $\underline{Y}'_2{}^{-1}$ from (20) requires $q^2(q^4 - 2q^3 + 3q^2 - q + 1)$ complex multiplications, where q is the number of accessible nodes in the network.

This method of computing $\underline{Y}'_2{}^{-1}$ is unnecessarily inefficient since many of the matrices in (17) and (20) are sparse, and others (\underline{S} and \underline{R}_{22}), in addition, are special since their non-zero elements are either +1 or -1. In Section 5 we show how to reduce the number of complex multiplications required to evaluate $\underline{Y}'_2{}^{-1}$ to $\beta q^2(q-1)$, where β is typically about 0.3 and is always a positive number less than unity.

Once $\underline{Y}'_2{}^{-1}$ is known, the y -parameters at the accessible nodes are found from (8). Evaluating the matrix products in (8) in a straightforward manner requires $2b^3 + (2p + q - 1)b^2 + (q-1)^2b$ multiplications, where

b is the number of branches in the network, p is the number of accessible nodes not counting the reference node, and q is the number of inaccessible nodes. Using the sparse matrix techniques described in Section 5, the number of multiplications needed to evaluate \underline{y}_1 from $\underline{y}_2'^{-1}$ can be reduced to $2(1 + \gamma)b^2$, where γ is typically about 0.1. For example, a network with 20 branches, 3 accessible nodes (plus the accessible reference node), and 8 inaccessible nodes requires 14180 multiplications to evaluate \underline{y}_1 directly from (8). By using sparse matrix techniques, this is reduced to 880 multiplications, a saving of a factor of 16 in this case. The straightforward evaluation in general requires nearly b times as many multiplications as the more efficient approach using sparse matrix techniques, so the saving increases with network complexity.

4. Open Circuits and Other Large Parameter Changes

For completeness, we conclude this paper with a discussion of open circuits and other large parameter changes, although this has been discussed elsewhere [4]. The main purpose in this section is to show how the analysis can be carried out in the context of nodal analysis.

Parameter changes are accounted for in (8) by a change in the $b \times b$ branch admittance matrix \underline{Y}_b . In order to evaluate the new y -parameters from (8), the inverse of the inaccessible nodal admittance matrix $\underline{Y}_2 = \underline{A}_2 \underline{Y}_b \underline{A}_2^T$ must be found. If a single parameter in the network is changed, then the new branch admittance matrix is $\underline{Y}_b' \triangleq \underline{Y}_b + \underline{\Delta}$, where $\underline{\Delta}$ has only one non-zero element representing the change in one parameter. Suppose a change occurs in element ij . Then $\underline{\Delta}$ can be written $d\underline{u}\underline{v}^T$, where d is a scalar equal to the parameter change, \underline{u} is a unit vector in direction i , and \underline{v} is a unit vector in direction j . Now define $\underline{U} \triangleq \underline{A}_2 \underline{u}$ and $\underline{V} = \underline{A}_2 \underline{v}$, so that $\underline{Y}_2' = \underline{Y}_2 + d\underline{U}\underline{V}^T$. Householder shows that the inverse of \underline{Y}_2' can be found from

$$\underline{Y}_2'^{-1} = \underline{Y}_2^{-1} [1 - \underline{U}(d^{-1} + \underline{V}^T \underline{Y}_2^{-1} \underline{U})^{-1} \underline{V}^T \underline{Y}_2^{-1}] \quad (21)$$

The inverse of \underline{Y}_2 is known, and $d^{-1} + \underline{V}^T \underline{Y}_2^{-1} \underline{U}$ is a scalar, so (21) can be evaluated without a matrix inversion.

5. Sparse Matrix Techniques

Sparse matrices arise in both the evaluation of the inverse of \underline{Y}'_2 and in the evaluation of the accessible y-parameter matrix \underline{Y}_1 . These are discussed separately in Sections 5.1 and 5.2. Section 5.3 summarizes the total saving in computational complexity and shows an example.

5.1 Evaluating the Inverse of \underline{Y}'_2

We have shown that the inverse of \underline{Y}'_2 may be found from (20), which is repeated below for convenience:

$$\underline{Y}'_2{}^{-1} = \underline{K}[1 + (\underline{R}_{22}\underline{Y}_2\underline{u})\underline{M}/(1 - \underline{M}(\underline{R}_{22}\underline{Y}_2\underline{u}))] \quad (21)$$

where \underline{K} and \underline{M} are defined by

$$\begin{bmatrix} \underline{K} & \underline{L} \\ \underline{M} & \underline{N} \end{bmatrix} = (\underline{S}^T)^{-1} \underline{Y}_2^{-1} \underline{S}^{-1} \quad (22)$$

The dimensions of \underline{K} and \underline{M} are respectively $(q-1) \times (q-1)$ and $1 \times (q-1)$.

In (22) \underline{Y}_2^{-1} is known, and \underline{S} is defined by

$$\underline{S} = \begin{bmatrix} \underline{R}_{22} \\ \underline{u}^T \end{bmatrix} \quad (23)$$

\underline{R}_{22} is the matrix which transforms \underline{A}_2 , the (inaccessible node)-to-branch incidence matrix before the fault, into \underline{A}'_2 , the (inaccessible node)-to-branch incidence matrix after the fault, so that

$$\underline{A}'_2 = \underline{R}_{22}\underline{A}_2 \quad (24)$$

In the discussion that follows, the reader may find it helpful to refer to the illustration of Fig. 3. Suppose nodes i and j are shorted, where $j > i$. Then rows 1 through $i-1$ of \underline{R}_{22} contain zeros except that the diagonal element is unity. This indicates that nodes 1 through $i-1$ in the faulted network are identical to nodes 1 through $i-1$ in the unfaulted network. Row i of \underline{R}_{22} contains all zeros except for unity elements in

columns i and j , indicating that node i in the faulted network is formed by combining nodes i and j in the unfaulted network. Finally, the remaining rows of \underline{R}_{22} contain all zeros, except that the first row to the right of the diagonal is unity. This indicates that the remaining nodes of the faulted network are identical to the remaining nodes of the unfaulted network. The faulted network contains one less node than the unfaulted network, so the dimension of \underline{R}_{22} is $(q-1) \times q$, where, as usual, q is the number of inaccessible nodes.

We have chosen to number the nodes consecutively, beginning with the accessible nodes. We have also avoided analysing networks with a short circuit between two accessible nodes, since such a fault may be found from driving point impedance measurements between all pairs of accessible nodes. Thus node j is always inaccessible, but node i may be either accessible or inaccessible. In the description above we have assumed that node i is inaccessible. If it is accessible, then \underline{R}_{22} is simpler than described above, since all rows contain a single unit element. (The unit element is on the diagonal in rows 1 through $j-1$ and in the first column to the right of the diagonal in all other rows. This indicates that all inaccessible nodes in the faulted network correspond to an inaccessible node in the unfaulted network, but one inaccessible node in the unfaulted network is now accessible, and so it has no row in \underline{R}_{22}).

In (23) \underline{u}^T is a row vector containing all zeros except that column j is unity.

To illustrate the formation of the \underline{S} matrix, consider a network with 8 inaccessible nodes. (For convenience in this discussion they are numbered 1 through 8, instead of $p+1$ through $p+8$). Figure 4(a) shows the \underline{S} matrix for a short circuit between nodes 4 and 6, and Fig. 4(b) shows the \underline{S} matrix for a short circuit between node 6 and any accessible node. The formation of the inverse of the \underline{S} matrix is given by $\underline{S}^{-1} = \underline{U}^T + \underline{Q}$, where

$$[\underline{Q}]_{rs} = \begin{cases} -1, & \text{if } r = i-p, s = n-p \\ 0, & \text{otherwise} \end{cases}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \textcircled{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

(a)

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

(b)

Fig. 4. \underline{S} matrices for a network with 8 inaccessible nodes. (a) Short circuit between nodes 4 and 6. (b) Short circuit between nodes 6 and any accessible node.

as explained in Section 3.2, where $\underline{S} = \underline{U} + \underline{P}$, and all elements of \underline{P} are zero except if the shorted nodes i and j are both inaccessible, in which case $[\underline{P}]_{ij}$ is unity. The unity element of \underline{P} is circled in Fig. 4(a). In this illustration we have taken $p=0$ for convenience, where p is the number of accessible nodes. Also note that $n-p = q$. The inverse of the \underline{S} matrix can be formed by inspection, and in particular, Fig. 5 shows the inverses of the matrices of Fig. 4.

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

(a)

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

(b)

Fig. 5. (a) Inverse of the matrix of Fig. 4(a). (b) Inverse of the matrix of Fig. 4(b)

We now turn to the problem of evaluating \underline{K} and \underline{M} . The right side of (22) may be written

$$(\underline{U} + \underline{Q}^T) \underline{Y}_2^{-1} (\underline{U}^T + \underline{Q}) = \underline{U} \underline{Y}_2^{-1} \underline{U}^T + \underline{U} \underline{Y}_2^{-1} \underline{Q} + \underline{Q}^T \underline{Y}_2^{-1} \underline{U}^T + \underline{Q}^T \underline{Y}_2^{-1} \underline{Q} \quad (23)$$

Recall that the rightmost column (column q) of (23) is not needed. The second and fourth terms on the right side of (23) are zero except for the rightmost column, so they need not be evaluated.

To evaluate $\underline{U} \underline{Y}_2^{-1} \underline{U}^T$, define the row vector $\underline{V}^T \triangleq [v_k]$, where $v_k =$ (unit column of row k in \underline{U}), $k = 1, 2, \dots, q-1$. Thus

$$[\underline{K}]_{mn} = [\underline{U} \underline{Y}_2^{-1} \underline{U}^T]_{mn} = [\underline{Y}_2^{-1}]_{v_m, v_n}, \quad \begin{cases} m = 1, 2, \dots, q-1 \\ n = 1, 2, \dots, q-1 \end{cases} \quad (24)$$

Finally \underline{M} is the $1 \times (q-1)$ row vector whose elements are the first $q-1$ columns of the last row (row q) of the sum of terms one and three of (23). Thus

$$\begin{aligned} [\underline{M}]_n &= [\underline{U} \underline{Y}_2^{-1} \underline{U}^T]_{q,n} + [\underline{Q}^T \underline{Y}_2^{-1} \underline{U}^T]_{q,n} \\ &= [\underline{Y}_2^{-1}]_{j, v_n} - [\underline{Y}_2^{-1}]_{i, v_n}, \quad (n = 1, 2, \dots, q-1) \end{aligned} \quad (25)$$

\underline{K} and \underline{M} are evaluated from (24) and (25) with no multiplications, and the evaluation of \underline{K} , the larger of the two, requires no additions.

Now that \underline{K} and \underline{M} are known, $\underline{Y}_2'^{-1}$ can be found from (21). In evaluating (21) it is useful to first find the $(q-1) \times 1$ column vector $\underline{x} \triangleq \underline{R}_{22} \underline{Y}_2 \underline{u}$. Recall that \underline{u} is a vector with all zero elements except that row j is unity. Thus \underline{x} is the j^{th} column of $\underline{R}_{22} \underline{Y}_2$. \underline{R}_{22} is the partition of \underline{S} formed from the first $q-1$ rows (23). We may make use of the row vector \underline{V}^T to form \underline{x} as follows:

$$[\underline{x}]_n = [\underline{R}_{22} \underline{Y}_2 \underline{u}]_n = \begin{cases} [\underline{Y}_2]_{v_n, j}, & \text{if node } i \text{ is accessible} \\ \left\{ \begin{aligned} &[\underline{Y}_2]_{v_n, j}, & v_n \neq i \\ &[\underline{Y}_2]_{v_n, j} + [\underline{Y}_2]_{j, j}, & v_n = i \end{aligned} \right\}, & \text{node } i \text{ is inaccessible} \end{cases} \quad (26)$$

($n = 1, 2, \dots, q-1$)

Using the definition of \underline{x} in (26), (21) may be written

$$\underline{Y}_2'^{-1} = \underline{K} + \underline{K}[\underline{xM}/(1 - \underline{Mx})] \quad (27)$$

which requires evaluating three matrix products, two of which directly involve \underline{x} . From (26) it is clear that \underline{x} is formed by reordering the elements of column j of \underline{Y}_2 . Since only a few inaccessible nodes are normally coupled to node j , \underline{x} is usually sparse. Define β equal to the fraction of non-zero elements of column j of \underline{Y}_2 . The number of multiplications required to compute \underline{Mx} is $\beta q \leq q$. β is usually significantly less than one, and so it is worth designing the algorithm to avoid the multiplications by zero. Similarly, βq rows of \underline{xM} are zero, and it is worth forming only the non-zero rows. Thus (27) can be evaluated with $\beta q^2(q-1)$ multiplications and divisions. Note that the multiplications in (27) are the only multiplications required to evaluate $\underline{Y}_2'^{-1}$, since \underline{K} and \underline{M} have been formed in (24) and (25) without multiplications.

5.2 Evaluating the y-Parameters

In this section we discuss the evaluation of the y-parameter matrix \underline{Y}_1 given by (8). In the faulted network, the node-to-branch incidence matrices are \underline{A}'_1 and \underline{A}'_2 , so that (8) becomes

$$\underline{Y}_1 = \underline{A}'_1 \underline{Y}_b [1 - \underline{A}'_2 \underline{Y}_2'^{-1} \underline{A}'_2 \underline{Y}_b] \underline{A}'_1^T \quad (28)$$

Two forms appear in (27) which must be discussed separately: $\underline{A}'^T \underline{Y}_A$ and $\underline{A} \underline{Y}_A^T$.

The most convenient way to store the node-to-branch incidence matrices, which we will refer to generally by \underline{A} in the discussion which follows, is by two column vectors \underline{A}^+ and \underline{A}^- , defined as follows. If the k^{th} column of \underline{A} has a +1 element, then $[\underline{A}^+]_k \triangleq$ (row number of the +1 element of the k^{th} column of \underline{A}), otherwise $[\underline{A}^+]_k \triangleq 0$. Similarly, if the k^{th} column of \underline{A} has a -1 element, then $[\underline{A}^-]_k \triangleq$ (row number of the -1 element of the k^{th} column of \underline{A}), otherwise $[\underline{A}^-]_k = 0$. The formation of \underline{A}^+ and \underline{A}^- corresponding to a given node-to-branch incidence matrix \underline{A} is illustrated in Fig. 6. Note that in practice it is necessary to store only the vectors \underline{A}^+ and \underline{A}^- and not the matrix \underline{A} .

$$\underline{A} = \begin{bmatrix} -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

$$\underline{A}^+ = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 2 \\ 3 \\ 3 \\ 4 \\ 0 \\ 5 \end{bmatrix} \quad \underline{A}^- = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \\ 5 \\ 4 \\ 4 \\ 0 \\ 5 \\ 0 \end{bmatrix}$$

Fig. 6. Node-to-branch incidence matrix \underline{A} , and the corresponding compact equivalent vectors \underline{A}^+ and \underline{A}^- .

Now consider the evaluation of $\underline{X} = \underline{A}^T \underline{Y} \underline{A}$. Element $[\underline{X}]_{m,n}$ is given by

$$\begin{aligned} [\underline{X}]_{m,n} &= \sum_k [\underline{A}^T]_{m,k} \sum_l [\underline{Y}]_{k,l} \cdot [\underline{A}]_{l,n} \\ &= \sum_{k,l} [\underline{A}]_{k,m} \cdot [\underline{A}]_{l,n} \cdot [\underline{Y}]_{k,l} \\ &= [\underline{Y}]_{[\underline{A}^+]_m, [\underline{A}^+]_n} + [\underline{Y}]_{[\underline{A}^-]_m, [\underline{A}^-]_n} \\ &\quad - [\underline{Y}]_{[\underline{A}^+]_m, [\underline{A}^-]_n} - [\underline{Y}]_{[\underline{A}^-]_m, [\underline{A}^+]_n} \end{aligned} \quad (29)$$

If any of the subscripts of $[\underline{Y}]$ in the last expression is zero, the term is zero.

The evaluation of $\underline{X}' = \underline{A} \underline{Y} \underline{A}^T$ is somewhat different because the summation is along the rows of \underline{A} instead of down the columns. In particular, element $[\underline{X}']_{m,n}$ is given by

$$\begin{aligned}
[\underline{X}']_{m,n} &= \sum_k [\underline{A}]_{m,k} \sum_l [\underline{Y}]_{k,l} \cdot [\underline{A}^T]_{l,n} \\
&= \sum_{k,l} [\underline{A}]_{m,k} \cdot [\underline{A}]_{n,l} \cdot [\underline{Y}]_{k,l}
\end{aligned}
\tag{30}$$

The right side of (30) cannot be evaluated by choosing m and n as in (29). Instead k and l are chosen and, for each choice, $[\underline{Y}]_{k,l}$ is added to elements $[\underline{X}']_{[\underline{A}^+]_k, [\underline{A}^+]_l}$ and $[\underline{X}']_{[\underline{A}^-]_k, [\underline{A}^-]_l}$ and subtracted from elements $[\underline{X}']_{[\underline{A}^+]_k, [\underline{A}^-]_l}$ and $[\underline{X}']_{[\underline{A}^-]_k, [\underline{A}^+]_l}$. If any of the subscripts of $[\underline{X}']$ is zero, that term is zero.

Products of the forms $\underline{X}' = \underline{A} \underline{Y} \underline{A}^T$ and $\underline{X} = \underline{A}^T \underline{Y} \underline{A}$ both appear in (28), and we have shown how to evaluate them without multiplication. An additional product of the form $\underline{X}' \cdot \underline{Y}_b$ must be evaluated, where \underline{Y}_b is given by (3). \underline{Y}_b is typically very sparse, so in the software we have chosen to store it as three vectors: one containing the non-zero elements of \underline{Y}_b , and the other two containing respectively the row and column of that element. The saving in storage can be appreciated by recognizing that \underline{Y}_b is a $b \times b$ matrix, where b is the number of branches in the network. The elements of \underline{Y}_b which are not zero are the b diagonal elements plus one additional element for each controlled source. If the number of controlled sources is $b/10$, then the sparse storage of \underline{Y}_b requires storing $1.1b$ complex numbers and $2.2b$ integers, compared with the b^2 complex numbers required for conventional storage. If the software is written to accommodate networks with 100 branches, then conventional storage requires setting aside room for 10^4 complex numbers, compared to about 120 complex numbers and 240 integers required by compact storage.

Considerable computation time is also saved by storing \underline{Y}_b sparsely. The matrix product $\underline{X}' \cdot \underline{Y}_b$ in (28) would require b^3 complex multiplications if performed directly, since the dimensions of both \underline{X}' and \underline{Y}_b are $b \times b$. The algorithm of Fig. 7 forms the product in $(1 + \gamma)b^2$ complex multiplications, where γ is the fraction of the b branches containing a controlled source. Typically γ is about 0.1. In Fig. 7, $YB(\cdot)$ is a vector containing the IYB non-zero elements of \underline{Y}_b , and $YBROW(\cdot)$ and $YBCOL(\cdot)$ are integer vectors containing the row and column numbers of each element of $YB(\cdot)$.

The product $\underline{X}' \cdot \underline{Y}_b$ is called $D(\cdot, \cdot)$. $D(\cdot, \cdot)$ is not sparse, so it is stored conventionally.

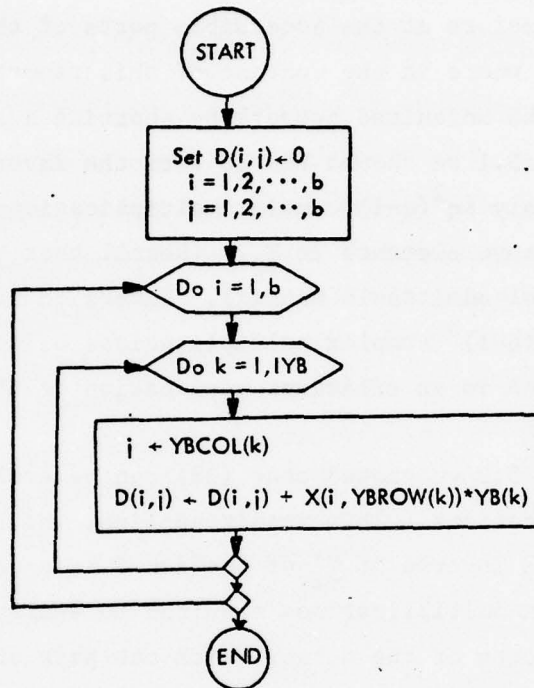


Fig. 7. Efficient formation of the matrix product $\underline{X}' \cdot \underline{Y}_b$.

Examination of (28) reveals another matrix product of the form $\underline{Y}_b \cdot \underline{Z}$, where in this case $\underline{Z} = \underline{1} + \underline{A}_2^T \underline{Y}_1^{-1} \underline{A}_2' \underline{Y}_b$. This matrix product is formed in a similar manner to $\underline{X}' \cdot \underline{Y}_b$ in $(1 + \gamma)b^2$ complex multiplications. Thus altogether $2(1 + \gamma)b^2$ complex multiplications are required to formulate \underline{y}_1 . A direct evaluation of (28) requires $2b^3 + (q-1+p)b^2 + ((q-1)^2 + p^2)b$ multiplications. Taking $p=2$, $b=20$, $q=11$ we find that \underline{y}_1 requires 22880 multiplications by direct methods, but only 880 multiplications by the methods described in this section with $\gamma = 0.1$, a saving of a factor of 26 in this example.

Before closing this section it should be noted that the y-parameters of the unfaulted network are evaluated before evaluating the y-parameters of any faulted networks. This is done by applying (28) with \underline{A}_1' replaced by \underline{A}_1 , \underline{A}_2' replaced by \underline{A}_2 , and \underline{Y}_2^{-1} replaced by \underline{Y}_2 . The inverse of \underline{Y}_2 must be

computed by conventional methods, but the methods of this section may be used for the remaining computation.

5.3 Summary of Efficient Computation

The y-parameters at the accessible ports of the faulted network are found from (28), where in the context of this report the faulted network is formed from the unfaulted network by shorting a single pair of nodes.

In Section 5.1 we showed how to form the inverse of the $(q-1) \times (q-1)$ matrix \underline{Y}'_2 with only $\beta q^2(q-1)$ complex multiplications, where β is the fraction of non-zero elements in \underline{Y}_2 . (Recall that $\underline{Y}_2 = \underline{A}_2 \underline{Y}_b \underline{A}_2^T$ is the inaccessible nodal admittance matrix). Inversion of a $(q-1) \times (q-1)$ matrix requires about $5(q-1)^3$ complex multiplications using Wilf's rank annihilation method¹[11], which is an efficient application of the Householder formula (19), [5-7].

In Section 5.2 we showed that (28) can be evaluated in $2(1 + \gamma)b^2$ complex multiplications. This result applies, whether or not the efficient evaluation of the inverse of \underline{Y}'_2 of section 5.1 is used. Thus the total number of complex multiplications required to evaluate the y-parameters at the accessible ports of the network with one pair of shorted terminals is $2(1 + \gamma)b^2 + q^2(q-1)$. If the inverse of \underline{Y}'_2 is found by conventional methods but the sparse matrix techniques of Section 5.2 are used to form \underline{y}_1 , then $2(1 + \gamma)b^2 + 5(q-1)^3$ complex multiplications are required.

The y-parameters resulting from all possible single short circuits involving at least one inaccessible node must be evaluated. If the network contains p accessible terminals and q inaccessible terminals, then it can be shown that $N = q(p + q) - q(q-1)/2$ distinct pairs of shorted terminals can be identified. The total number of complex multiplications is therefore $N[2(1 + \gamma)b^2 + \beta q^2(q-1)]$ using the efficient method in inverting \underline{Y}'_2 and $N[2(1 + \gamma)b^2 + 5(q-1)^3]$ using a conventional inversion method. To show the benefit of the efficient matrix inversion, let us assume that the number of accessible ports, p, is 2, the number of branches b is twice the

¹ An error appears in our edition of Ralston and Wilf [11]. In the flow chart on p. 75, step 11 should be 1→i, not 2→i.

number of nodes, $p + q + 1$, the fraction γ of branches having controlled sources is 0.1, and the fraction β of non-zero elements of \underline{Y}'_2 is 0.3. Using these assumptions, the total number of complex multiplications necessary to analyze all the inaccessible short circuits using both the efficient and the conventional methods of inverting \underline{Y}'_2 are shown in Fig. 8. The advantage of the efficient method is especially noticable for the larger networks.

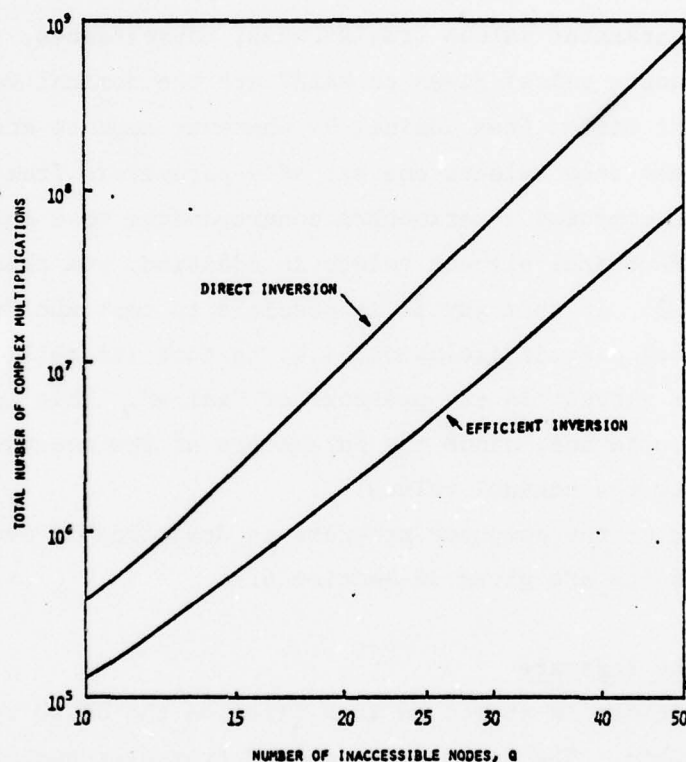


Fig. 8. Number of complex multiplications required to analyze all inaccessible short circuits in a network with q inaccessible nodes.

6. Status of Current Software

The algorithms described above have been implemented in Fortran IV. Currently two main programs are available which access the same sub-programs. Each accept as data a description of the network, including the identification of the accessible nodes. MAIN1 then prints the y-

parameters at the accessible terminals computed from the nominal network and from every network that can be obtained from the nominal network by shorting pairs of terminals in which at least one terminal in the pair is inaccessible. MAIN2 is identical to MAIN1 except that in addition it requires as data a set of y-parameters measured at the accessible terminals. MAIN2 then compares all computed y-parameters to the measured y-parameters, determines the closest match, and takes this case as the most likely short circuit.

MAIN1 may be used to generate test data for MAIN2. Used in this way the network parameter values (resistances, capacitances, inductances, and controlled source gains) given to MAIN2 are the nominal values, but those given to MAIN1 differ from nominal by whatever amounts are desired for the test. The user then selects one set of y-parameters from the printout of MAIN1. These computed y-parameters, corresponding to a specific short circuit with off-nominal element values in addition, are then used as "measured" data for MAIN2. In this way it is possible to test the "robustness" of the method of short circuit isolation; i.e. to test its ability to find the correct short circuit in the presence of "noise". This is, of course, necessary in practice, since the parameters of the measured network will not be exactly the nominal values.

The use of the computer programs is described in Section 6.1, and some test results are given in Section 6.2.

6.1 Using the Software

The software is stored on disk files on the DEC10 system in Building 620, WPAFB, Ohio. The files are accessed from user number [7777,137], password "ALFIE". The Fortran (*.FOR) files are stored under the names MAIN1.FOR, MAIN2.FOR, BLOCK4.FOR, FAULT4.FOR, SNEW4.FOR, and MINVC.FOR. To create a load module to execute MAIN1, the following command is used:

```
.LOAD MAIN1,BLOCK4, FAULT4,SNEW4,MINVC
```

A load module to execute MAIN2 is formed in exactly the same way, except that "MAIN1" is replaced by "MAIN2". (The .SAVE command must be given

immediately after the .LOAD command, or the load module will not be stored on the disk).

In the interest of saving disk space, all relocatable binary (*.REL) files have been deleted. They may be created, if desired, by the .COM command. For example, .COM FAULT4.FOR will cause the FAULT4.FOR file to be compiled by the Fortran compiler and stored under the name FAULT4.REL.

The load modules have already been created and stored under the names MAIN1.EXE and MAIN2.EXE. To execute these files from the CRT terminal, the command .RUN MAIN1 or .RUN MAIN2 is used.

After the command .RUN MAIN1 has been given, the following prompt message appears on the screen:

```
ENTER IP, N, FREQ                                     (31)
```

The dots under the text are guides marking the beginning and end of each of the three data fields. (The input format is 2I5, E10.5). The first data value (IP) is an integer equal to the number of accessible nodes (not counting the reference node, node 0). The second data value (N) is an integer equal to the total number of nodes in the network (not counting the reference node). The third data value (FREQ) is a real number equal to the frequency in Hz at which the analysis is to be performed.

After entering these three numbers (and pressing the "return" key), the following prompt message appears on the screen:

```
ENTER NB, N1, N2, ITYPE, VALUE                       (32)
```

where again the dots are guides to the beginning and end of each data field. (In this case the input format is 4I5, E10.5). This message prompts the user to enter the description of a network branch or controlled source. In order to understand how to enter this data, some conventions must be explained. Each network branch contains a single passive element (resistor, capacitor, or inductor). The controlled sources are associated with passive branch elements and are not counted as branches. Branches are

numbered consecutively, beginning with 1. Otherwise the branch numbering is arbitrary, and the branch data need not be entered in same order as the branch numbers. Each branch is connected between two nodes. The nodes are numbered consecutively beginning with 0. By the convention we have chosen, node 0 is the reference node for the node voltages. The accessible nodes are numbered 1, 2, ..., IP. The software is dimensioned to accept at most 3 accessible nodes ($IP \leq 3$), 17 inaccessible nodes ($N \leq 20$), and 50 branches. In entering data for a passive branch, and referring to (32), the first data value (NB) is an integer equal to the branch number, the second and third data values (N1 and N2) are integers equal to the node numbers at the two ends of the branch, the fourth data value (ITYPE) is an integer indicating the type of passive element (1 = resistor, 2 = inductor, 3 = capacitor), and the fifth data value (VALUE) is the element value (in ohms, henries, or farads, as appropriate). The order in which the two node numbers (N1 and N2) are entered establish the branch reference direction, as shown in Fig. 9. This is important only if a controlled



Fig. 9. Reference directions for the branch current j_k and branch voltage v_k for branch k.

source is coupled to or from the branch. If no controlled source is coupled to or from the branch, N1 may be either node number and N2 the other. Controlled source data will be discussed in the next paragraph. The prompt message (32) appears on the screen each time the return key is pressed. After all branch data is entered (including controlled source data), enter 99999 followed by a "return". The prompt message "NETWORK INPUT COMPLETE" will then appear on the screen to reassure the user that all is well so far.

Controlled source data is entered under the same prompt message (32) as the passive branch data and may be mixed in with the passive branch data or entered last, as the user prefers. In entering controlled source data, the first data value entered (NB) is a five digit inter beginning with 4, such as 40001. The same number may be used for all controlled sources. The second and third data values (N1 and N2) are the branch numbers of the "to" and "from" branches respectively. The "to" branch is the branch containing the source, and the "from" branch is the branch which controls the source. Fig. 10 shows the conventions for the four types of controlled sources (1 = transconductance, 2 = transresistance, 3 = current gain, 4 = voltage gain). The fourth data value entered is an integer equal to the type of controlled source, and the fifth data value entered is a real number equal to the gain value (in mhos for type 1 sources, ohms for type 2 sources, or dimensionless for type 3 and 4 sources).

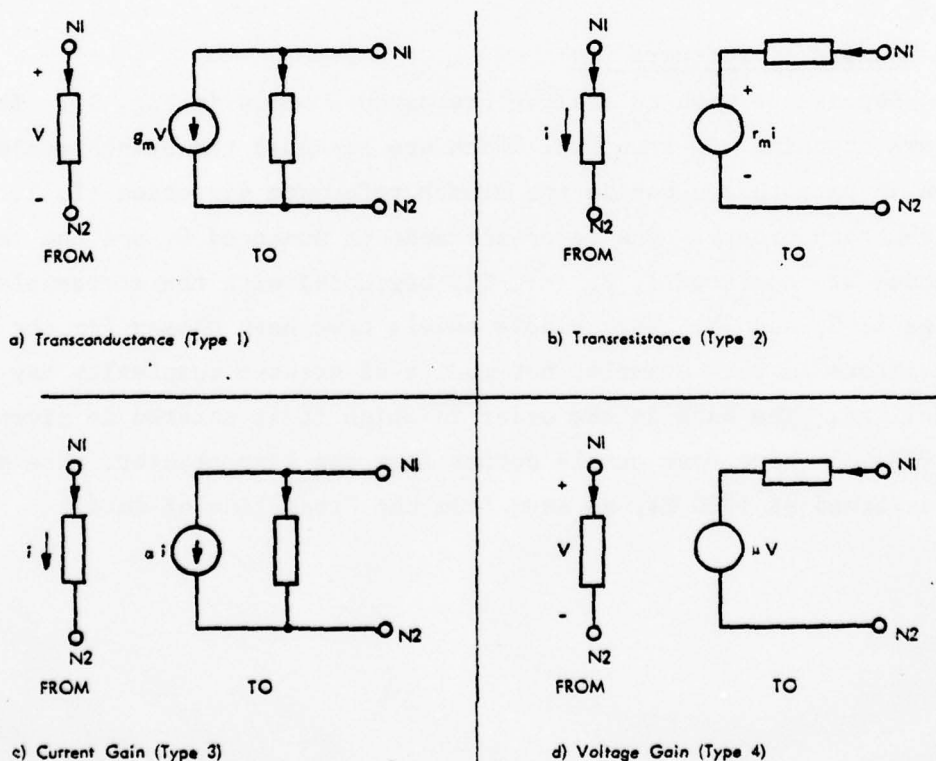


Fig. 10. The four types of controlled sources, showing their branch conventions.

This completes the data requested if MAIN1 is executed. If MAIN2 is executed, the prompt message below appears after "NETWORK INPUT COMPLETE":

ENTER YN(1,1) (33)
.
.
.

The dots below the text indicate the beginning the end of the two data fields, which in this case are the real and imaginary parts respectively of the measured y-parameter in mhos. The number of measured y-parameters requested depends on the number of accessible ports (1 if one port is accessible, 4 if two ports are accessible, and 9 if three ports are accessible).

The major output from the program appears on the line printer.

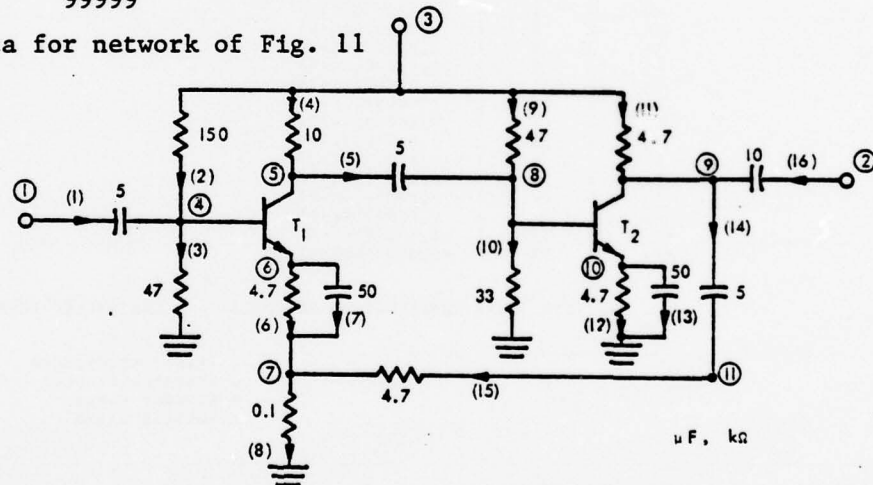
An example illustrating the use of the program is given in Section 6.2.

6.2 Example of Software Use

Suppose we wish to analyse the network shown in Fig. 11. This network contains 20 branches, which are assigned the branch numbers shown in parentheses beside the branch reference direction (1, 2, ..., 20 in arbitrary order). The reference node is numbered 0, and the remaining 11 nodes are numbered 1, 2, ..., 11, beginning with the accessible nodes (nodes 1, 2, and 3). Very simple models have been chosen for the two transistors in this example, but models of greater complexity may be used if desired. The data in the order in which it is entered is given in Fig.12, and Fig. 13 shows some sample output from the line printer. The analysis is performed at 1000 Hz, as seen from the first line of data.

3	11	1000.		
1	1	4	3	5.E-6
2	3	4	1	150.E3
3	4	0	1	47.E3
4	3	5	1	10.E3
5	5	8	3	5.E-6
6	6	7	1	4.7E3
7	6	7	3	50.E-6
8	7	0	1	100.
9	3	8	1	47.E3
10	8	0	1	33.E3
11	3	9	1	4.7E3
12	10	0	1	4.7E3
13	10	0	3	50.E-6
14	9	11	3	5.E-6
15	11	7	1	4.7E3
16	9	2	3	10.E-6
17	4	6	1	2.E3
18	5	6	1	50.E3
19	8	10	1	2.E3
20	9	10	1	50.E3
40001	18	17	1	0.05
40002	20	19	1	0.05
99999				

Fig. 12. Data for network of Fig. 11



μ F, kΩ

Ground is node 0

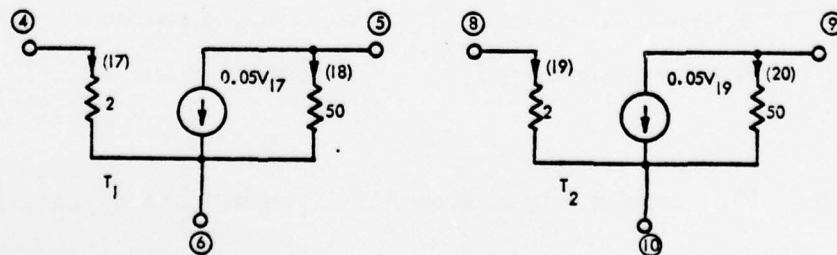


Fig. 11. Feedback amplifier.

NETWORK DATA:

11 NODES
3 ACCESSIBLE NODES

ANALYSIS AT 0.1000E+02 HZ

NETWORK DATA CARDS

1	1	4	3	0.4500000E+01
2	3	4	1	0.1400000E+06
3	4	0	1	0.5500000E+05
4	3	5	1	0.1100000E+05
5	5	8	3	0.5500000E+01
6	6	7	1	0.4200000E+04
7	6	7	3	0.5500000E+04
8	7	0	1	0.9400000E+02
9	3	4	1	0.5200000E+05
10	8	0	1	0.3000000E+05
11	3	5	1	0.5100000E+04
12	10	0	1	0.4400000E+04
13	10	0	3	0.5100000E+04
14	9	11	3	0.5400000E+05
15	11	7	1	0.4600000E+04
16	9	2	3	0.1090000E+04
17	4	6	1	0.2100000E+04
18	5	6	1	0.5400000E+05
19	8	10	1	0.1900000E+04
20	9	10	1	0.4800000E+05
**	18	17	1	-0.4000000E+01
**	20	19	1	-0.4500000E+01

NETWORK INPUT COMPLETE: 20 BRANCHES, 2 CONTROLLED SOURCES INCLUDING:

2 TRANSCONDUCTANCES
0 TRANSRESISTANCES
0 CURRENT GAINS
0 VOLTAGE GAINS

NOMINAL Y-PARAMETERS, MHUF

0.5556E-04	0.2317E-04 J	0.6087E-06	0.3313E-06 J	-0.8679E-05	-0.1809E-05 J
0.1050E-01	-0.5673E-01 J	-0.6117E-03	0.5653E-03 J	0.1962E-02	0.1139E-02 J
-0.1637E-01	-0.3752E-02 J	-0.6318E-04	0.1496E-03 J	0.6295E-03	-0.4600E-03 J

Fig. 13. Sample output from MAIN1, using data of Fig. 12.

Y-PARAMETERS WITH NODES 0 AND 11 SHORTED

0.8908E-04	-0.4213E-04 J	-0.2099E-09	-0.1433E-09 J	-0.6980E-05	0.5706E-07 J
0.4859E-01	0.7708E-02 J	0.8628E-04	0.2406E-03 J	-0.8267E-03	0.1425E-02 J
0.2284E-02	-0.1139E-01 J	-0.1271E-03	-0.2464E-04 J	0.6414E-03	0.2155E-03 J

Y-PARAMETERS WITH NODES 1 AND 11 SHORTED

0.2390E-01	0.3868E-02 J	0.4274E-04	-0.2201E-03 J	-0.4166E-03	0.7060E-03 J
0.4766E-01	0.7333E-02 J	0.8628E-04	0.2406E-03 J	-0.8267E-03	0.1425E-02 J
0.2175E-02	-0.1117E-01 J	-0.1271E-03	-0.2464E-04 J	0.6414E-03	0.2155E-03 J

Y-PARAMETERS WITH NODES 2 AND 11 SHORTED

0.8908E-04	-0.4213E-04 J	0.2251E-06	0.8736E-06 J	-0.6980E-05	0.5706E-07 J
0.7265E-01	0.1146E-01 J	-0.1048E-02	-0.1988E-03 J	-0.1236E-02	0.2131E-02 J
0.2284E-02	-0.1139E-01 J	-0.2359E-03	0.1911E-03 J	0.6414E-03	0.2155E-03 J

Fig. 13, Concluded

6.3 Effectiveness of Software in Locating Short Circuit Faults

The network of Fig. 11 was used to test the effectiveness of the software (MAIN2) in locating a single short circuit. The nominal element values for the network were entered in addition to the 9 complex y-parameters at the three available ports (nodes 1-0, 2-0, and 3-0) "measured" from the faulty network. The "measured" data was generated from MAIN1 using element values set from 10% to 20% above or below the nominal values. Table 1 shows the actual fault, the fault determined by the minimum cost function, and the cost function of the actual fault. The cost function used was

$$c = \sum_{i,j} \frac{|y_{ij} - \hat{y}_{ij}|}{|\hat{y}_{ij}|} \quad (34)$$

where y_{ij} is the y-parameter computed by MAIN2 and \hat{y}_{ij} is the measured y-parameter (in this case computed from MAIN1). Of the seven examples at 1000 Hz, the program correctly located five of the faults. The two errors seemed to be caused by the fact that the magnitude of the capacitor impedances at 1000 Hz were much lower than the impedance levels of the network resistances. (The magnitude of the impedance of a 10 μ F capacitor

Table 1. Summary of Test Results

FREQUENCY HZ	ACTUAL SHORTED NODES	SHORTED NODES OF MINIMUM COST FUNCTION	MINIMUM COST FUNCTION	COST FUNCTION OF CORRECT FAULT
1000	0-8	0-8	1.087	1.087
	0-5	0-5	1.115	1.115
	0-10	0-10	1.290	1.290
	6-7	5-8	1.271	1.34
	3-5	3-8	1.122	1.20
	3-9	3-9	2.828	2.828
	7-8	7-8	1.632	1.632
10	6-7	6-11	3.928	5.08
	3-5	3-8	2.290	2.36

at 1000 Hz is 15.9Ω). The test was rerun at 10 Hz for the two incorrectly identified cases, and the faults were still not correctly identified. It was expected that the short circuit between nodes 6 and 7 would be difficult to identify, since the $50 \mu\text{F}$ bypass capacitor is only about 3Ω at 1000 Hz, and so the changes in the parameters of the measured circuit could be expected to mask the short circuit. In fact, one of the reasons for running the tests was to see how much the network parameters could be changed before this change prevented reliable identification of the short circuit. Variations of 10% to 20% still permitted correct identification in 5 out of 7, or 71% of the cases tried. If the parameters of the "measured" network were set to the nominal values, then correct identification would occur, of course, in 100% of the cases, since then no error would exist between the "measured" y-parameters and those generated by MAIN2 for the correct fault. A sequence of tests should be conducted with increasing amounts of off-nominal parameter variation. This has not yet been done due to time limitations.

III CONCLUSIONS

We have shown an efficient way to analyse networks containing a single short circuit or other larger parameter change. The efficiency results from the fact that the network change caused by either the short circuit or other large parameter change results in an alteration of the inaccessible nodal admittance matrix by a matrix of unit rank. The response of the faulted network can therefore be computed in terms of the nominal response with many fewer multiplications than are required for and original analysis. Sparse matrix techniques were also used to reduce execution time.

The efficiency of the algorithm makes it practical to exhaustively search all possible single short circuits involving an inaccessible node in order to find a solution to the problem of locating a short circuit in an analog network from measurements at the accessible terminals. The test examples described in Section 6.3 show that short circuits can be reliably located by the algorithm, even though measurements were made at only one frequency and differences between the actual and the nominal parameter values were not taken into account. We expect that the reliability can be improved by making measurements at more than one frequency and by solving for the parameter values.

IV RECOMMENDATIONS

The algorithm described in this report is an important first step in the development of a practical algorithm for fault analysis in analog circuits. The test results are most encouraging and justify further work to improve the reliability of the method.

Some steps which should be taken to improve the reliability of the fault isolation algorithm are summarized below.

- Investigate the problem of determining a set of frequencies at which independent measurements can be made. The more independent data that is available, the more successful the fault isolation can be expected to be.

- Investigate the efficacy of incorporating parameter (resistance, capacitance, inductance, controlled source gain) changes into the fault model in addition to short circuits. We have discussed an efficient way to recomputing the y-parameters at the accessible nodes after the occurrence of a single large parameter change, but this is not implemented in the present algorithm. The investigation should determine if a multi-dimensional search for the best parameter fit improves the reliability of the algorithm, and if it can be done fast enough to justify its inclusion in the algorithm.

APPENDIX A
COMPUTER PROGRAM LISTINGS

The listings of all Fortran computer programs described or referred to in this report are listed in this appendix. These programs are stored in disk files under user number [7777,137], password "ALFIE", on the DEC10 System in Building 620, WPAFB, Ohio. The programs are listed under the various file names in which they are stored: MAIN1, MAIN2, BLOCK4, ELEMNT, OTHER4, FAULT4, SNEW, and MINVC.

Two non-standard statements (INCLUDE ' . ') appear in the listings: INCLUDE 'OTHER4.FOR' and INCLUDE 'ELEMNT.FOR'. These statements cause the statements in files OTHER4.FOR or ELEMNT.FOR to be inserted in place of the INCLUDE statement. The INCLUDE statements have been used to insert COMMON statements in the appropriate places.

AD-A065 650

OHIO STATE UNIV RESEARCH FOUNDATION COLUMBUS
USAF-ASEE (1978) SUMMER FACULTY RESEARCH PROGRAM (WPAFB). VOLUM--ETC(U)
NOV 78 C D BAILEY

F/G 1/3

F44620-76-C-0052

UNCLASSIFIED

AFOSR-TR-79-0231

NL

6 OF 6

AD
A065650



END
DATE
FILMED
5-79
DDC

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

BLOCK4.FOR

```
00100      BLOCK DATA
00300      INCLUDE 'OTHER4.FOR'
00500      DATA A2PLUS,A2MNUS/100*0/
00600      DATA A1PLUS,A1MNUS/100*0/
00700      END
```

ELEMNT.FOR

```
00100      COMMON/ELEMNT/Y(65), GM(15), RM(15), XMU(15), ALPHA(15),
00200      1      YB(65), YROW(65), GMROW(15), GMCOL(15),
00300      2      RMROW(15), RMCOL(15), XMUROW(15), XMUCOL(15),
00400      3      AROW(15), ACOL(15), YBROW(65), YBCOL(65), IGM,
00500      4      IRM, IXMU, IA, IALPHA, IYB
00600      INTEGER YROW, GMROW, GMCOL, RMROW, RMCOL, XMUROW, XMUCOL,
00700      1      AROW, ACOL, YBROW, YBCOL
00800      COMPLEX Y, YB
```

OTHER4.FOR

```
00100      COMMON/OTHER/A1PLUS(50), A1MNUS(50), A1PLSP(50), A1MNSP(50),
00200      1      Y2(17,17), Y2I(17,17), Y2PI(16,16), A2PLUS(50),
00300      2      A2MNUS(50), A2PLSP(50), A2MNSP(50), I, J, IP,
00400      3      N, IB, ITOT, YN(3,3)
00500      INTEGER A2PLUS, A2MNUS, A2PLSP, A2MNSP
00550      INTEGER A1PLUS, A1MNUS, A1PLSP, A1MNSP
00600      COMPLEX YN, Y2, Y2I, Y2PI
```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

MAIN1.FOR

```

00020 C      THIS PROGRAM COMPUTES THE Y-PARAMETERS OF AN ELECTRICAL
00040 C      NETWORK WITH INTERNAL SHORT CIRCUITS.  AN EFFICIENT
00060 C      ALGORITHM IS USED.  SEE REFERENCE:
00080 C
00085 C      A.T. JOHNSON, JR., 'EFFICIENT FAULT ANALYSIS IN ANALOG
00090 C      CIRCUITS', FINAL REPORT, USAF-ASEE SUMMER FACULTY
00095 C      RESEARCH PROGRAM, WPAFB, OHIO, AUGUST 1978
00098 C
00100 C      INCLUDE 'ELEMNT.FOR'
00200 C      INCLUDE 'OTHER4.FOR'
00210 C
00230 C      READ NETWORK DATA AND COMPUTE NODE-TO-BRANCH INCIDENCE
00250 C      MATRICES (SPARSE STORAGE) OF ORIGINAL NETWORK: A1PLUS,
00270 C      A1MNUS, A2PLUS, A2MNUS
00300 C      CALL INPUT
00330 C
00350 C      COMPUTE YB MATRIX: REF., EQ.(3).
00370 C
00400 C      CALL YBMAT
00500 C      IQ = N-IP
00550 C
00570 C      COMPUTE Y2 = A2 * YB * A2T.  REF., EQ.(2).
00590 C
00600 C      CALL Y2PMAT(Y2,IQ,17,A2PLUS,A2MNUS)
00650 C
00690 C      COMPUTE INVERSE OF Y2.
00695 C
00700 C      CALL MINVC(Y2,Y2I,IQ,17)
00750 C
00770 C      COMPUTE Y-PARAMETERS AT ACCESSIBLE TERMINALS.
00790 C      REF., EQ.(8).
00795 C
00800 C      CALL YMAT(A1PLUS,A1MNUS,A2PLUS,A2MNUS,Y2I,YN,IP,IB,17)
00900 C      WRITE(3,102)
01000 C      DO 15 M=1,IP
01100 15  WRITE(3,100)(YN(M,NX),NX=1,IP)
01200 C      NCASES = IQ*N - ((IQ-1)*IQ)/2
01230 C
01250 C      LOOP THROUGH ALL PAIRS OF SHORT CIRCUITS INVOLVING
01270 C      INACCESSIBLE NODES.  I AND J ARE THE SHORTED NODES.
01290 C
01300 C      DO 10 K = IP+1,N
01400 C      WRITE(5,110) NCASES
01500 110  FORMAT(/5X,I3,' CASES LEFT')
01600 C      J = N + IP + 1 - K
01700 C      NCASES = NCASES - J
01800 C      DO 10 L=1,J
01900 C      I = L-1

```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO

```
01920 C
01940 C      COMPUTE NEW NODE-TO-BRANCH INCIDENCE MATRICES (A1PLSP,
01960 C      A1MNSP, A2PLSP, A2MNSP) WITH NODES I AND J SHORTED.
01980 C      REF., EQ.(20).
01985 C
02000      CALL ANEW
02020 C
02040 C      COMPUTE INVERSE OF Y2P FOR FAULTED NETWORK.
02060 C      REF., EQ.(28).
02080 C
02100      CALL SNEW
02200      CALL YMAT(A1PLSP,A1MNSP,A2PLSP,A2MNSP,Y2PI,YN,IP,IB,16)
02300      WRITE(3,101) I,J
02330 C
02350 C      PRINT Y-PARAMETERS.
02370 C
02400      DO 10 M=1,IF
02500 10  WRITE(3,100) (YN(M,NX),NX=1,IP)
02600      STOP
02700 100  FORMAT( 3(10X, 2E12.4, ' J') //)
02800 101  FORMAT('1',10X,'Y-PARAMETERS WITH NODES',I3,
02900 *   ' AND',I3,' SHORTED'////)
03000 102  FORMAT(///,10X,'NOMINAL Y-PARAMETERS, MHOS: '//)
03100      END
```


MAIN2.FOR

```

00100 C
00200 C      THIS PROGRAM ISOLATES SHORT CIRCUIT FAULTS IN ANALOG NETWORKS.
00300 C      THE FAULT IS ISOLATED BY COMPARING COMPUTED Y-PARAMETERS AT
00400 C      THE ACCESSIBLE NODES (FOR THE NETWORK WITH A PAIR OF NODES
00500 C      SHORTED) TO MEASURED Y-PARAMETERS AT THE SAME ACCESSIBLE
00600 C      NODES. FOR A DESCRIPTION OF THE EFFICIENT ALGORITHM, SEE:
00700 C
00800 C      A.T. JOHNSON, JR., 'EFFICIENT FAULT ANALYSIS IN ANALOG
00900 C      CIRCUITS', FINAL REPORT, USAF-ASEE SUMMER FACULTY RESEARCH
01000 C      PROGRAM, WPAFB, OHIO, AUGUST 1978
01100 C
01200 C      INCLUDE 'ELEMNT.FOR'
01300 C      INCLUDE 'OTHER4.FOR'
01400 C      DIMENSION DENOM(3,3)
01500 C      COMPLEX YNNOM(3,3)
01600 C
01700 C      READ NETWORK DATA AND COMPUTE NODE-TO-BRANCH INCIDENCE MATRICES
01800 C      (SPARSE STORAGE) OF ORIGINAL NETWORK, A1PLUS, A1MNUS, A2PLUS,
01900 C      A2MNUS.
02000 C
02100 C      CALL INPUT
02200 C
02300 C      READ MEASURED Y-PARAMETERS: YNNOM
02400 C
02500 C      DO 5 K=1,IP
02600 C      DO 5 L=1,IP
02700 C      WRITE(5,106) K,L
02800 C      WRITE(5,110)
02900 C      READ(5,108) YNNOM(K,L)
03000 C      DENOM(K,L) = CABS(YNNOM(K,L))
03100 5 IF(DENOM(K,L) .EQ. 0.) DENOM(K,L) = 1.E-9
03200 C      WRITE(3,109)
03300 C      DO 1 M=1,IP
03400 1 WRITE(3,100)(YNNOM(M,NX),NX=1,IP)
03500 C      IQ= N-IP
03600 C
03700 C      COMPUTE YB MATRIX. SEE REF., EQ.(3).
03800 C
03900 C      CALL YBMAT
04000 C
04100 C      COMPUTE Y2 = A2 * YB * A2T. SEE REF., EQ.(2).
04200 C
04300 C      CALL Y2PMAT(Y2,IQ,17,A2PLUS,A2MNUS)
04400 C
04500 C      COMPUTE INVERSE OF Y2.
04600 C
04700 C      CALL MINVC(Y2,Y2I,IQ,17)
04800 C
04900 C      COMPUTE Y-PARAMETERS AT ACCESSIBLE TERMINALS. REF., EQ.(8).
05000 C
05100 C      CALL YMAT(A1PLUS,A1MNUS,A2PLUS,A2MNUS,Y2I,YN,IP,IB,17)
05200 C      ISAVE = 0
05300 C      JSAVE = 0

```

```

05400 C
05500 C PRINT Y-PARAMETERS OF UNFAULTED NETWORK WITH NOMINAL
05600 C PARAMETER VALUES.
05700 C
05800 C WRITE(3,102)
05900 C DO 2 M=1,IP
06000 2 WRITE(3,100)(YN(M,NX),NX=1,IP)
06100 C
06200 C COMPUTE COST FUNCTION OF UNFAULTED NETWORK. CALL IT COSTMN.
06300 C
06400 C COSTMN = 0.
06500 C DO 3 I1=1,IP
06600 C DO 3 J1=1,IP
06700 3 COSTMN = COSTMN + CABS(YN(I1,J1) - YNNOM(I1,J1))/DENOM(I1,J1)
06800 C WRITE(3,107) COSTMN
06900 C NCASES = IQ*N - ((IQ-1)*IQ)/2
07000 C
07100 C LOOP THROUGH ALL PAIRS OF SHORT CIRCUITS INVOLVING AN
07200 C INACCESSIBLE NODE. I AND J ARE THE SHORTED NODES.
07300 C
07400 C DO 10 K = IP+1,N
07500 C WRITE(5,111) NCASES
07600 111 FORMAT(/5X,I3,' CASES LEFT')
07700 C J = N + IP + 1 - K
07800 C NCASES = NCASES - J
07900 C DO 10 L=1,J
08000 C I = L-1
08100 C
08200 C COMPUTE NEW NODE-TO-BRANCH INCIDENCE MATRICES (A1PLSP, A1MNSP,
08300 C A2PLSP, A2MNSP) WITH NODES I AND J SHORTED. SEE REF., EQ. (13).
08400 C
08500 C CALL ANEW
08600 C
08700 C COMPUTE INVERSE OF Y2P FOR FAULTED NETWORK. SEE REF., EQ.(10).
08800 C
08900 C CALL SNEW
09000 C
09100 C COMPUTE Y-PARAMETERS AT ACCESSIBLE TERMINALS OF FAULTED
09200 C NETWORK. SEE REF., EQ.(28).
09300 C
09400 C CALL YMAT(A1PLSP,A1MNSP,A2PLSP,A2MNSP,Y2PI,YN,IP,IB,16)
09500 C
09600 C COMPUTE COST FUNCTION OF CURRENT Y-PARAMETERS (NODES I AND J
09700 C SHORTED).
09800 C
09900 C COST = 0.
10000 C DO 20 I1=1,IP
10100 C DO 20 J1=1,IP
10200 20 COST = COST + CABS(YN(I1,J1) - YNNOM(I1,J1))/DENOM(I1,J1)
10300 C IF(COST - COSTMN) 22,21,21
10400 22 COSTMN = COST
10500 C ISAVE = I
10600 C JSAVE = J

```

```

10700 C
10800 C PRINT NODES I AND J AND THE COMPUTED Y-PARAMETERS AT THE
10900 C ACCESSIBLE TERMINALS.
11000 C
11100 21 WRITE(3,101) I,J
11200 DO 30 M=1,IP
11300 30 WRITE(3,100) (YN(M,NX),NX=1,IP)
11400 10 WRITE(3,107) COST
11500 C
11600 C PRINT NODES OF LOWEST COST FAULT AND COST FUNCTION FOR
11700 C THIS PAIR OF SHORTED NODES.
11800 C
11900 WRITE(3,103) ISAVE, JSAVE, COSTMN
12000 STOP
12100 100 FORMAT (3(10X,2E12.4,' J')//)
12200 101 FORMAT('1',10X,'Y-PARAMETERS WITH NODES',I3,
12300 * ' AND',I3,' SHORTED'//)
12400 102 FORMAT(///,10X,'BEGIN SEARCH USING NOMINAL NETWORK.'//,
12500 * 10X,'Y-PARAMETERS WITH NO SHORT CIRCUITS, MHOS: '//)
12600 103 FORMAT('1',10X,'MOST LIKELY SINGLE SHORT: '//11X,'NODES',
12700 * I3,' AND',I4,' SHORTED.'//10X,'(COST FUNCTION =',
12800 * E11.4,')' )
12900 106 FORMAT(///,5X,'ENTER YN(',I1,',',I1,')' )
13000 107 FORMAT(///,56X,'COST =', E11.3)
13100 108 FORMAT(E10.4,1X,E10.4)
13200 109 FORMAT(///,10X,'MEASURED Y-PARAMETERS, MHOS. '//)
13300 110 FORMAT(' . . . ')
13400 200 FORMAT(8(F8.2,F6.2,' J'))
13500 END

```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO

```

00100      SUBROUTINE INPUT
00200      C
00300      C      THIS SUBROUTINE READS THE NETWORK DATA AND STORES THE NODE-
00400      C      TO-BRANCH INCIDENCE MATRICES.
00500      C
00600      C      VARIABLE DEFINITIONS:
00700      C
00800      C      IB      - NUMBER OF BRANCHES IN NETWORK (PASSIVE ELEMENTS)
00900      C      IGM      - NUMBER OF TRANSCONDUCTANCES
01000      C      IRM      - NUMBER OF TRANSRESISTANCES
01100      C      IYMU      - NUMBER OF VOLTAGE GAINS
01200      C      IALPHA    - NUMBER OF CURRENT GAINS
01300      C      N        - NUMBER OF NODES (NOT COUNTING REFERENCE NODE)
01400      C      IP        - NUMBER OF ACCESSIBLE NODES ( ' ' )
01500      C      FREQ      - FREQUENCY AT WHICH ANALYSIS IS PERFORMED (HZ)
01600      C      Y        - MATRIX OF PASSIVE ELEMENT ADMITTANCES (SPARSE)
01700      C      GM        - MATRIX OF TRANSCONDUCTANCES (SPARSE)
01800      C      RM        - MATRIX OF TRANSRESISTANCES (SPARSE)
01900      C      XMU       - MATRIX OF VOLTAGE GAINS (SPARSE)
02000      C      ALPHA     - MATRIX OF CURRENT GAINS (SPARSE)
02100      C      YROW(.)    VECTOR STORING ROW OF Y(.)
02200      C      GMROW(.)   VECTOR STORING ROW OF GM(.)
02300      C      GMCOL(.)   VECTOR STORING COLUMN OF GM(.)
02400      C
02500      C      SIMILARLY FOR RMROW, RMCOL, XMUROW, XMUCOL, AROW, ACOL.
02600      C
02700      C      INCLUDE 'OTHER4.FOR'
02800      C      INCLUDE 'ELEMNT.FOR'
02900      C      WRITE(5,103)
03000      103  FORMAT(10('/'),' ENTER IP, N, FREQ')
03100      C      WRITE(5,108)
03200      108  FORMAT(' . . . . . ')
03300      C      READ(5,100) IP,N,FREQ
03400      C      OMEGA = 2.*3.141593*FREQ
03500      C      IB = 0
03600      C      IGM = 0
03700      C      IRM = 0
03800      C      IXMU = 0
03900      C      IALPHA = 0
04000      C      WRITE(3,102) N, IP, FREQ
04100      1    WRITE(5,104)
04200      104  FORMAT(5('/'),' ENTER NB, N1, N2, ITYPE, VALUE')
04300      C      WRITE(5,109)
04400      109  FORMAT(' . . . . . ')
04500      C      READ(5,101)NB,N1,N2,ITYPE,VALUE
04600      C      IF(NB .EQ. 99999) GO TO 200
04700      C      WRITE(3,106)NB,N1,N2,ITYPE,VALUE
04800      C      IF (NB - 10000) 10,20,20
04900      10   IB = IB+1
05000      C      IF (N1.EQ.0) GO TO 15
05100      C      IF (N1-IP) 16,16,17
05200      16   A1PLUS(NB) = N1
05300      C      GO TO 15
05400      17   A2PLUS(NB) = N1-IP
05500      15   IF(N2.EQ.0) GO TO 18
05600      C      IF(N2-IP) 19,19,30
05700      19   A1MNUS(NB) = N2
05800      C      GO TO 18
05900      30   A2MNUS(NB) = N2-IP
06000      18   GO TO (11,12,13) ITYPE

```



```

06100 11 Y(IB) = (1.,0.)/CMPLX(VALUE,0.)
06200 GO TO 2
06300 12 Y(IB) = (0.,-1.)/(OMEGA*VALUE)
06400 GO TO 2
06500 13 Y(IB) = (0.,1.)*OMEGA*VALUE
06600 2 YROW(IB) = NB
06700 GO TO 1
06800 20 GO TO (21,22,23,24) ITYPE
06900 21 IGM = IGM + 1
07000 GM(IGM) = VALUE
07100 GMROW(IGM) = N1
07200 GMCOL(IGM) = N2
07300 GO TO 1
07400 22 IRM = IRM + 1
07500 RM(IRM) = VALUE
07600 RMROW(IRM) = N1
07700 RMCOL(IRM) = N2
07800 GO TO 1
07900 23 IALPHA = IALPHA + 1
08000 ALPHA(IALPHA) = VALUE
08100 AROW(IALPHA) = N1
08200 ACOL(IALPHA) = N2
08300 GO TO 1
08400 24 IXMU = IXMU + 1
08500 XMU(IXMU) = VALUE
08600 XMUROW(IXMU) = N1
08700 XMUCOL(IXMU) = N2
08800 GO TO 1
08900 200 ITOT = IGM + IRM + IALPHA + IXMU
09000 WRITE(5,110)
09100 WRITE(3,105) IB, ITOT
09200 WRITE(3,107) IGM, IRM, IALPHA, IXMU
09300 RETURN
09400 100 FORMAT (2I5,E10.5)
09500 101 FORMAT (4I5,E10.5)
09600 102 FORMAT ('1',25X,'NETWORK DATA: '//16X,I3,' NODES '//16X,I3,' ACCESSIBL
09700 *E NODES '//16X,' ANALYSIS AT',E12.4,' HZ',10(/),2X,' NETWORK DATA CAR
09800 *DS'//)
09900 105 FORMAT (//25X,' NETWORK INPUT COMPLETE:',I3,' BRANCHES,',
10000 1 I4,' CONTROLLED SOURCES INCLUDING: '//)
10100 106 FORMAT(20X,I3,3I5,E16.7)
10200 107 FORMAT(I60,' TRANSCONDUCTANCES '//I60,' TRANSRESISTANCES '//
10300 1 I60,' CURRENT GAINS '//I60,' VOLTAGE GAINS')
10400 110 FORMAT('// NETWORK INPUT COMPLETE')
10500 END

```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DQC

```

10600      SUBROUTINE YBMAT
10700      C
10800      C      THIS SUBROUTINE COMPUTES MATRIX YB.  SEE REF., EQ.(3)
10900      C
11000      INCLUDE 'OTHER4.FOR'
11100      INCLUDE 'ELEMNT.FOR'
11200      INTEGER XROW(80), XCOL(80), WROW(80), WCOL(80)
11300      COMPLEX X(80), W(80)
11400      C
11500      C      FORM X = -Y*XMU.  ALL MATRICES ARE STORED SPARSELY.
11600      C      SINCE Y IS 'DIAGONAL', COMPUTATION IS SIMPLIFIED.
11700      C
11800      IF(IXMU .EQ. 0) GO TO 11
11900      DO 10 L = 1, IXMU
12000          I1 = XMURROW(L)
12100          XROW(L) = I1
12200          XCOL(L) = XMUCOL(L)
12300          DO 9 K=1, IB
12400              IF(YROW(K) .EQ. I1) GO TO 10
12500          9      CONTINUE
12600          10      X(L) = -Y(K)*XMU(L)
12700          11      IX = IXMU
12800      C
12900      C      FORM W = -Y*RM.  ALL MATRICES ARE STORED SPARSELY.
13000      C      SINCE Y IS 'DIAGONAL', COMPUTATION IS SIMPLIFIED.
13100      C
13200      IF(IRM .EQ. 0) GO TO 13
13300      DO 12 L=1, IRM
13400          I1 = RMROW(L)
13500          WROW(L) = I1
13600          WCOL(L) = RMCOL(L)
13700          DO 14 K=1, IB
13800              IF(YROW(K) .EQ. I1) GO TO 12
13900          14      CONTINUE
14000          12      W(L) = -Y(K)*RM(L)
14100          13      IW = IRM
14200      C
14300      C      FORM X = Y + X + GM.  ALL MATRICES ARE STORED SPARSELY.
14400      C      SINCE Y IS 'DIAGONAL', AND X AND GM HAVE NEITHER DIAGONAL
14500      C      ELEMENTS NOR ELEMENTS IN COMMON, COMPUTATION IS SIMPLIFIED.
14600      C
14700      DO 20 L=1, IB
14800          X(IX+L) = Y(L)
14900          XROW(IX+L) = YROW(L)
15000          20      XCOL(IX+L) = YROW(L)
15100          IX = IX + IB
15200      IF(IGM .EQ. 0) GO TO 31
15300      DO 30 L=1, IGM
15400          X(IX+L) = GM(L)
15500          XROW(IX+L) = GMROW(L)
15600          30      XCOL(IX+L) = GMCOL(L)
15700          31      IX = IX + IGM

```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

```

15800 C
15900 C FORM W = 1 + W + ALPHA. ALL MATRICES ARE STORED SPARSELY.
16000 C SINCE W AND ALPHA HAVE NO COMMON ELEMENTS AND NO DIAGONAL
16100 C ELEMENTS, COMPUTATION IS SIMPLIFIED.
16200 C
16300 DO 40 L=1,IB
16400 W(IW+L) = (1.,0.)
16500 WROW(IW+L) = L
16600 40 WCOL(IW+L) = L
16700 IW = IW + IB
16800 IF(IALPHA .EQ. 0) GO TO 51
16900 DO 50 L=1,IALPHA
17000 W(IW+L) = -ALPHA(L)
17100 WROW(IW+L) = AROW(L)
17200 50 WCOL(IW+L) = ACOL(L)
17300 51 IW = IW + IALPHA
17400 C
17500 C FORM YB = W*X. ALL MATRICES ARE STORED SPARSELY.
17600 C COMPUTATION IS COMPLICATED BY THE FACT THAT EXISTING
17700 C ELEMENTS MUST BE SOUGHT BEFORE THE MULTIPLICATION
17800 C CAN BE PERFORMED.
17900 C
18000 IYB = 0
18100 DO 90 K=1,IB
18200 DO 80 L=1,IW
18300 IF(WCOL(L) - K) 80,55,80
18400 55 DO 70 M=1,IX
18500 IF(XROW(M) - K) 70,65,70
18600 65 IF(IYB .EQ. 0) GO TO 67
18700 DO 60 N1=1,IYB
18800 IF(YBROW(N1) - WROW(L)) 60,75,60
18900 75 IF(YBCOL(N1) - XCOL(M)) 60,85,60
19000 60 CONTINUE
19100 67 IYB = IYB + 1
19200 YB(IYB) = W(L)*X(M)
19300 YBROW(IYB) = WROW(L)
19400 YBCOL(IYB) = XCOL(M)
19500 GO TO 70
19600 85 YB(N1) = YB(N1) + W(L)*X(M)
19700 70 CONTINUE
19800 80 CONTINUE
19900 90 CONTINUE
20000 RETURN
20100 END

```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO

```

20200      SUBROUTINE ANEW
20300      C
20400      C      THIS SUBROUTINE COMPUTES THE NODE-TO-BRANCH INCIDENCE MATRICES
20500      C      (SPARSELY STORED VECTORS A1PLSP, A1MNSP, A2PLSP, A2MNSP) FOR
20600      C      THE NETWORK WITH NODES I AND J SHORTED.  SEE REF., EQ.(13).
20700      C
20800      INCLUDE 'OTHER4.FOR'
20900      DO 150 L=1,IB
21000      A1PLSP(L) = A1PLUS(L)
21100      A1MNSP(L) = A1MNUS(L)
21200      IF(A2PLUS(L) - (J-IP)) 20,30,40
21300      20  A2PLSP(L) = A2PLUS(L)
21400      GO TO 90
21500      30  IF(I-IP) 50,50,60
21600      50  IF(I.EQ. 0) GO TO 55
21700      IF(A1MNSP(L) - I) 52,51,52
21800      51  A1MNSP(L) = 0
21900      GO TO 55
22000      52  A1PLSP(L) = I
22100      55  A2PLSP(L) = 0
22200      GO TO 90
22300      60  IF(A2MNUS(L) - (I-IP)) 80,70,80
22400      70  A2PLSP(L) = 0
22500      A2MNSP(L) = 0
22600      GO TO 90
22700      80  A2PLSP(L) = I-IP
22800      GO TO 90
22900      40  A2PLSP(L) = A2PLUS(L) - 1
23000      90  IF(A2MNUS(L) - (J-IP)) 100,110,120
23100      100 A2MNSP(L) = A2MNUS(L)
23200      GO TO 150
23300      110 IF(I-IP) 130,130,135
23400      130 IF(I.EQ. 0) GO TO 134
23500      IF(A1PLSP(L) - I) 132,131,132
23600      131 A1PLSP(L) = 0
23700      132 A1MNSP(L) = I
23800      134 A2MNSP(L) = 0
23900      GO TO 150
24000      135 IF(A2PLUS(L) - (I-IP)) 145,140,145
24100      140 A2PLSP(L) = 0
24200      A2MNSP(L) = 0
24300      GO TO 150
24400      145 A2MNSP(L) = I-IP
24500      GO TO 150
24600      120 A2MNSP(L) = A2MNUS(L) - 1
24700      150 CONTINUE
24800      RETURN
24900      END

```


THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

```

25000 SUBROUTINE YMAT(A1PLSP,A1MNSP,A2PLSP,A2MNSP,Y2PI,YN,IPP,IB,IQT)
25100 C
25200 C THIS SUBROUTINE COMPUTES THE Y-PARAMETERS AT THE ACCESSIBLE
25300 C NODES. SEE REF., EQ.(8). SPARSE MATRIX TECHNIQUES ARE
25400 C USED, AS DESCRIBED IN REF., PP. 18-22.
25500 C
25600 INCLUDE 'ELEMNT.FOR'
25700 COMPLEX Y2PI(IQT,IQT),YN(3,3)
25800 INTEGER A2PLSP(50),A2MNSP(50)
25900 INTEGER A1PLSP(50),A1MNSP(50)
26000 COMPLEX W,C(50,50),D(50,50),E(50,50)
26100 C
26200 DO 100 I=1,IB
26300 I1 = A2PLSP(I)
26400 I2 = A2MNSP(I)
26500 DO 100 J=1,IB
26600 D(I,J) = (0.,0.)
26700 E(I,J) = (0.,0.)
26800 J1 = A2PLSP(J)
26900 J2 = A2MNSP(J)
27000 W = (0.,0.)
27100 IF(I1*J1 .NE. 0) W = W + Y2PI(I1,J1)
27200 IF(I2*J2 .NE. 0) W = W + Y2PI(I2,J2)
27300 IF(I1*J2 .NE. 0) W = W - Y2PI(I1,J2)
27400 IF(I2*J1 .NE. 0) W = W - Y2PI(I2,J1)
27500 100 C(I,J) = W
27600 C
27700 DO 5 I=1,IB
27800 DO 5 L=1,IYB
27900 K = YBROW(L)
28000 J = YBCOL(L)
28100 5 D(I,J) = D(I,J) - C(I,K)*YB(L)
28200 C
28300 DO 10 K=1,IB
28400 10 D(K,K) = D(K,K) + (1.,0.)
28500 C
28600 DO 15 J=1,IB
28700 DO 15 L=1,IYB
28800 I = YBROW(L)
28900 K = YBCOL(L)
29000 15 E(I,J) = E(I,J) + YB(L)*D(K,J)
29100 C
29200 DO 20 I=1,IPP
29300 DO 20 J=1,IPP
29400 20 YN(I,J) = (0.,0.)
29500 C
29600 C COMPUTE A1 * E * A1T. A1 IS STORED SPARSELY.
29700 C
29800 DO 30 L=1,IB
29900 IP = A1PLSP(L)
30000 IM = A1MNSP(L)
30100 DO 30 K=1,IB
30200 JP = A1PLSP(K)
30300 JM = A1MNSP(K)
30400 IF(IP*JP .NE. 0) YN(IP,JP) = YN(IP,JP) + E(L,K)
30500 IF(IM*JM .NE. 0) YN(IM,JM) = YN(IM,JM) + E(L,K)
30600 IF(IP*JM .NE. 0) YN(IP,JM) = YN(IP,JM) - E(L,K)
30700 IF(IM*JP .NE. 0) YN(IM,JP) = YN(IM,JP) - E(L,K)
30800 30 CONTINUE
30900 RETURN
31000 END

```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDQ

```
31100 SUBROUTINE Y2PHAT(Y2P,IQP,IQPT,A2PLSP,A2MNSP)
31200 C
31300 C THIS SUBROUTINE FORMS EITHER Y2 = A2 * YB * A2T OR
31400 C Y2P = A2P * YB * A2PT. SEE REF., EQ.(2). SPARSE MATRIX
31500 C TECHNIQUES HAVE BEEN USED TO STORE YB.
31600 C
31700 C INCLUDE 'ELEMNT.FOR'
31800 C COMPLEX Y2P(IQPT,IQPT)
31900 C INTEGER A2PLSP(50), A2MNSP(50)
32000 C
32100 C INITIALIZE Y2P TO 0.
32200 C
32300 DO 10 I=1,IQP
32400 DO 10 J=1,IQP
32500 10 Y2P(I,J) = (0.,0.)
32600 C
32700 C COMPUTE Y2P = A2 * YB * A2T
32800 C
32900 DO 100 M=1,IYB
33000 L = YBROW(M)
33100 K = YBCOL(M)
33200 IP = A2PLSP(L)
33300 IM = A2MNSP(L)
33400 JP = A2PLSP(K)
33500 JM = A2MNSP(K)
33600 IF(IP*JP .NE. 0) Y2P(IP,JP) = Y2P(IP,JP) + YB(M)
33700 IF(IM*JM .NE. 0) Y2P(IM,JM) = Y2P(IM,JM) + YB(M)
33800 IF(IP*JM .NE. 0) Y2P(IP,JM) = Y2P(IP,JM) - YB(M)
33900 IF(IM*JP .NE. 0) Y2P(IM,JP) = Y2P(IM,JP) - YB(M)
34000 100 CONTINUE
34100 RETURN
34200 END
```

SNEW4.FOR

```

00100      SUBROUTINE SNEW
00110      C
00115      C      THIS SUBROUTINE FORMS THE INVERSE OF Y2P, CALLED Y2PI,
00120      C      WITH NODES I AND J SHORTED, USING EQ.(20) IN THE REF.
00125      C      SPARSE MATRIX TECHNIQUES ARE USED.  SEE REF., PP. 14-18.
00130      C
00200      INCLUDE 'OTHER4.FOR'
00500      COMPLEX XK(16,16),XM(16),C(16),XI(16,16),W,MC
00700      INTEGER V(17), XIROW(16)
00720      C
00740      C      FORM VECTOR V.  SEE REF., P. 17
00760      C
00900      DO 10 K=1,17
01000      10  V(K) = 0
01100      IF(J .EQ. IP+1) GO TO 25
01200      DO 20 K=1,J-IP-1
01300      20  V(K) = K
01400      25  IQ = N-IP
01500      DO 30 K = J-IP,IQ-1
01600      30  V(K) = K+1
01700      V(IQ) = J-IP
01800      C
01820      C      FORM MATRICES K AND M, DEFINED IN REF., EQ.(22).  SPARSE
01840      C      MATRIX TECHNIQUES ARE USED.  SEE REF., EQ.(24) AND (25).
01860      C
02200      DO 50 IR=1,IQ
02300      DO 50 IS=1,IQ-1
02400      W = Y2I(V(IR),V(IS))
02500      IF(IR-IQ) 60,65,65
02600      60  XK(IR,IS) = W
02700      GO TO 50
02800      65  XM(IS) = W-Y2I(I-IP,V(IS))
02900      50  CONTINUE
02920      C
02940      C      FORM R22 * Y2 * U.  SEE REF., EQ.(26).
02960      C
03000      DO 70 M=1,IQ-1
03100      70  C(M) = Y2(V(M),J-IP)
03400      IF(I .GT. IP) C(I-IP) = C(I-IP) + Y2(J-IP,J-IP)
03420      C
03440      C      FORM 1 - M * (R22 * Y2 * U).  SEE REF., EQ.(21).
03460      C
03500      MC = (1.,0.)
03600      DO 80 M=1,IQ-1
03650      IF(C(M) .NE. (0.,0.)) MC = MC - XM(M)*C(M)
03750      80  CONTINUE
03800      C
03810      C      FORM XI = C * XM / MC.  SINCE C HAS MANY ZERO ELEMENTS,
03820      C      XI HAS MANY ZERO ROWS.  ONLY THE NON-ZERO ROWS OF XI ARE
03830      C      FORMED BELOW.  SEE REF., EQ. (21).
03840      C

```

**THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDO**

```
03850      IROW = 0
03860      DO 85 K=1,IQ-1
03870          IF(C(K).EQ.(0.,0.)) GO TO 85
03880          IROW = IROW + 1
03890          XIROW(IROW) = K
03900          DO 85 M=1,IQ-1
03910              XI(IROW,M) = C(K)*XM(M)/MC
03920      85  CONTINUE
03930      C
03940      C      FORM Y2PI = XK * (XI + 1).  RECALL THAT ONLY THE NON-
03950      C      ZERO ROWS OF XI HAVE BEEN SAVED.  SEE REF., EQ.(21).
03960      C
03970      DO 95 I1=1,IQ-1
03980      DO 95 J1=1,IQ-1
03990          W = (0.,0.)
04000          DO 90 L=1,IROW
04010              K = XIROW(L)
04020      90  W = W + XK(I1,K)*XI(L,J1)
04030      95  Y2PI(I1,J1) = W + XK(I1,J1)
04500      RETURN
04600      END
```


MINVC.FOR

```

00100      SUBROUTINE MINVC(B,C,NN,NT)
00200      C
00300      C      THIS SUBROUTINE INVERTS THE COMPLEX VALUED MATRIX B.
00400      C      IN THE CALLING PROGRAM, B HAS BEEN DIMENSIONED NT X NT.
00500      C      THE ACTUAL DIMENSION USED IN THE CALLING PROGRAM IS NN X NN.
00600      C      THE INVERSE OF B IS RETURNED AS C, WHICH IS DIMENSIONED
00700      C      AS B.
00800      C
00900      COMPLEX ALPHA(17),BETA(17),SUM,SUMP,U(17),V(17),LAMBDA
01000      COMPLEX A(17,17),B(NT,NT),C(NT,NT)
01100      DO 1 I=1,NN
01200      DO 1 J=1,NN
01300      C(I,J) = (0.,0)
01400      1 A(I,J) = B(I,J)
01500      DO 3 I=1,NN
01600      A(I,I) = A(I,I) - (1.,0.)
01700      3 C(I,I) = (1.,0.)
01800      DO 40 N = 1,NN
01900      DO 10 I = N+1,NN
02000      ALPHA(I) = A(I,N)/A(N,N)
02100      10 BETA(I) = A(N,I)
02200      DO 17 I=1,NN
02300      SUM = (0.,0.)
02400      SUMP = (0.,0.)
02500      DO 15 K=N,NN
02600      SUM = SUM + C(I,K)*A(K,N)
02700      15 SUMP = SUMP + A(N,K)*C(K,I)
02800      U(I) = SUM
02900      17 V(I) = SUMP
03000      SUM = (0.,0.)
03100      DO 21 J = N,NN
03200      21 SUM = SUM + V(J)*A(J,N)
03300      LAMBDA = SUM + A(N,N)
03400      DO 30 I = 1,NN
03500      DO 30 J = 1,NN
03600      30 C(I,J) = C(I,J) - U(I)*V(J)/LAMBDA
03700      IF(N .EQ. NN) RETURN
03800      DO 38 I = N+1,NN
03900      DO 38 J = N+1,NN
04000      38 A(I,J) = A(I,J) - ALPHA(I)*BETA(J)
04100      40 CONTINUE
04200      END

```

References

- [1] C.A. Desoer and E.S. Kuh, Basic Circuit Theory. New York: McGraw-Hill, 1969, pp. 409-463.
- [2] A.T. Johnson, Jr. and A.J. Pennington, "Network equations with coupling elements", Proc. of the 14th Midwest Symposium on Circuit Theory, pp. 2.5-1 - 2.5-7, May 1971.
- [3] M.N. Ransom and R. Saeks, "Fault Isolation with Insufficient Measurements," IEE Trans. on Circuit Theory, Vol. CT-20, No. 4, pp. 416-417, July 1973.
- [4] H.S.M. Chen and R. Saeks, "Research on Fault Analysis", Annual Research report, Texas Tech, Lubbock, TX, 1977.
- [5] A.S. Householder, "A survey of some closed methods for inverting matrices," J. SIAM, Vol. 5, pp. 155-169, 1957.
- [6] B. Noble, Applied Linear Algebra, Englewood Cliffs, NJ Prentice Hall, 1969, pp. 147-148.
- [7] Ibid., p. 313.
- [8] V.N. Faddeeva, Computational Methods of Linear Algebra, New York: Dover Publications, 1959, pp. 102-103.
- [9] N. Sen and R. Saeks, "A Measure of Testability and its Application to Test Point Selection - Theory", Proc. of the 20th Midwest Symposium on Circuits and Systems, Texas Tech. Univ., Lubbock, Texas, August 1977, pp. 576-583.
- [10] N. Sen and R. Saeks, "A Measure of Testability and its Application to Test Point Selection - Computation", Proc of AUTOTESTCON '77. pp. 212-219, Hyannis, Mass., Nov. 1977.
- [11] A. Ralston and H. Wilf, Mathematical Methods for Digital Computers, Vol.1, John Wiley: New York 1967, pp. 73-77.
- [12] R. Saeks and S.R. Liberty, Rational Fault Analysis, New York: Marcel Dekker, Inc., 1977
- [13] N. Sen and R. Saeks, "A Measure of Testability and its Application to Test Point Selection - Theory", Proc. of the 20th Midwest Symposium on Circuits and Systems, Texas Tech U, Aug. 1977, pp. 576-583

- [14] N. Sen and R. Saeks, "A Measure of Testability and its Application to Test Point Selection - Computation", Proc. AUTOTESCON '77, Hyannis, Mass, Nov. 1977, pp. 212-219
- [15] "Large ATE Systems", Eval. Eng., vol. 17, no. 2, March/April 1978, pp. 28-50
- [16] E.A. Torrero, "ATE - Not So Easy", IEEE Spectrum, vol. 14, no. 4, April 1977, pp. 29-34
- [17] W.A. Pllice, "Techniques for the Automatic Generation of Fault Isolation Tests for Analog Circuits", AES Newsletter, Aut. 1976, pp. 27-30

ELECTRICAL PROPERTIES OF
MAGNESIUM AND GERMANIUM IMPLANTED
GALLIUM ARSENIDE

Frank L. Pedrotti
Avionics Laboratory, WPAFB

August 8, 1978

ELECTRICAL PROPERTIES OF MAGNESIUM
AND GERMANIUM IMPLANTED GALLIUM ARSENIDE

ABSTRACT

Preliminary results have been obtained in a Hall/sheet resistivity study of the electrical properties of Germanium (Ge)- and Magnesium (Mg)-implanted Gallium Arsenide (GaAs). Germanium samples implanted with 120 KeV ions at room temperature and in six doses ranging from 5×10^{12} to 1×10^{15} ions/cm² were found to produce p-type activity after encapsulation with silicon nitride and anneal for 15 minutes at 900°C. Activation efficiencies over 25% and mobilities between 100 and 200 cm²/V-sec were determined. At this time optimum activation and mobility cannot be linked with confidence to implantation and annealing conditions. Although Ge-implanted GaAs is known to be amphoteric, p-type activity has not been previously reported to our knowledge. Concentration profiles of Mg-implanted samples as a function of depth were measured using Hall/sheet resistivity measurements and a chemical etchant to strip away measured uniform surface layers. Because of the rapid diffusion of Mg ions, encapsulation was with silicon dioxide, taking advantage of its low deposition temperature of 325°C compared to 700°C for silicon nitride. By varying anneal times at a given temperature, the measured diffusion profiles permit a determination of diffusion coefficients for Mg ions of varying doses in GaAs by fitting with a Gaussian distribution which includes diffusion rate and time. All encapsulations were done in a pyrolytic reactor.

Frank L. Pedrotti
Marquette University

This research project involved the characterization of ion-implanted gallium arsenide by electrical measurements. Ion species investigated include both Germanium and Magnesium. The research accomplished during the ten week period represents only a beginning to the work which is being planned along these lines by this laboratory. In accordance with the goals of this USAF sponsored program, some of this work will be continued at Marquette University in collaboration with scientists of this laboratory.

This work was supported by the Air Force Office of Scientific Research through the USAF-ASEE Summer Faculty Research Program (WPAFB), Contract F44620-76-C-0052, the Ohio State University, Columbus, Ohio.

INTRODUCTION

Gallium Arsenide (GaAs) has been the most thoroughly studied among compound semiconductors. Some of this work has led to the fabrication of laser and LED devices, microwave elements including IMPATT diodes and Gunn diodes, and others. There are also promising applications in the field of integrated optoelectronics.

There are several reasons for the use of ion-implantation as a means of doping. Especially in the case of donor impurities which have low diffusion coefficients, thermal equilibrium doping techniques require that the sample be subjected to high temperature processing which leads to compositional changes due to thermal instability of the compound. Also the material tends to resist type conversion due to autocompensation by intrinsic lattice defects. The technique of ion implantation, in which a beam of dopant ions is energetically driven into the substrate surface, offers the possibility of avoiding these difficulties because it is essentially a non-equilibrium process. However, this technique produces some radiation damage of the lattice which must be annealed away, again requiring elevated temperatures and the possibility of the recurrence of the problems mentioned above. When such effects can be kept small, however, by the techniques to be discussed presently, ion implantation offers a means of controlled doping of crystals for the fabrication of a variety of useful devices.

In the process of implantation, the ions must give up their kinetic energies to the host lattice, resulting in radiation damage. The lattice defects so introduced must then be annealed out by appropriate thermal treatment. Often such defects have disassociation energies high enough so as to require annealing temperatures close to the melting point, or in the neighborhood of 900-1000 °C. However, annealing of the GaAs at such temperatures leads to disassociation and out-diffusion of Ga and As, as well as the implanted dopant

and Chromium, when it is used as a substrate dopant to produce high resistivity. An effective means to prevent such out-diffusion appears to be a capping of the implanted sample with an appropriate material. The cap then functions as a barrier to diffusion, protecting the sample during the high temperature anneal. Improper encapsulation leads to low activation of the implanted species and anomalous carrier concentration changes in the substrate material, masking the effects of the implantation. A thin film barrier that functions in this way must itself be mechanically stable and able to withstand high temperatures without blistering or loss of adhesion. Because metallic films can produce serious intermetallic diffusion with the semiconductor, dielectrics have usually been used. Both silicon dioxide (SiO_2) and silicon nitride (Si_3N_4) have been used as encapsulants, although the latter is favored for n-type dopants because it has been shown that SiO_2 does not effectively prevent outdiffusion of gallium. The presence of gallium vacancies may actually facilitate the substitutional location of p-type dopants. Even when successful encapsulation is achieved, --that is, when the cap remains integral during the anneal and effectively prevents out-diffusion -- the anneal itself may cause changes in the substrate. Such changes may include redistribution of chromium, diffusion of the implanted dopant and diffusion of defects. The effect of the annealing must then be well known and taken into consideration in the use of this technique to fabricate device materials with specific properties.

The ability to produce doped semiconductor materials with properties tailored to a specific application requires a clear understanding of the final depth distribution of the imbedded ions (the profile), as well as the conditions under which these ions will move into substitutional sites in the host lattice to provide free carriers (activation) of reasonable mobility. The ideal of 100% activation efficiency occurs when the implanted atoms occupy substitutional sites,

freely ionized, and not compensated. The highest mobilities are achieved under these circumstances.

A wide range of experimental techniques have been utilized in order to determine the relevant properties (characterization) of ion implanted semiconductor materials. These methods include photoluminescence, cathodoluminescence, glow discharge optical spectroscopy, Hall effect, capacitance-voltage measurements, proton-bombardment-induced x-ray analysis, Auger analysis, ellipsometry, and others. The work being reported here involves characterization by the use of Hall effect/sheet resistivity measurements in conjunction with repeated removal of successive surface layers of the implanted sample by chemical etching in order to determine a profile of the implanted species. The work has concentrated on a study of carrier concentration, activation efficiency, and carrier mobility of the implanted sample as a function of implanted dose, anneal temperature and time, and depth penetration.

EXPERIMENTAL PROCEDURES

Sample Preparation

The procedure outlined in the following describes the treatment of all samples studied, from the cutting of the substrate to final measurements.

Substrate material is received in the form of thin wafers about .020 inch as sliced from the boule. These wafers were $\langle 100 \rangle$ - oriented GaAs single crystals made semiinsulating by doping with chromium, and obtained from Crystal Specialties, Inc. The wafer is cleaved after scribing with a diamond stylus into .2 x .2 inch squares. Each sample is then cleaned while anchored into position over a small hole providing suction. The cleaning sequence consisted of several solvents directed in turn onto the sample surface from squeeze bottles for 10-20 seconds each: 10% aquasol solution, deionized water, trichloroethylene, acetone, and methanol. The sample was blow-dried with a jet of nitrogen gas after the water and methanol wash. Each sample was subsequently free-etched in a teflon beaker by submerging in a solution of H_2SO_4 :30% H_2O_2 : H_2O in a ratio of 3:1:1 by volume for 90 seconds. The cleaning and etching were done immediately before ion implantation.

Ion implantation of the samples followed. An ion current of approximately 1 microamp and energy of 120 KeV was directed at the sample until doses ranging from 3×10^{12} to 1×10^{15} ions/cm² were absorbed. The installation includes a means of suppressing secondary electron emission in order to determine total doses accurately from ion current measurements. The samples were implanted at room temperature with the beam from a hot cathode source, directed 6-10 degrees off from the $\langle 100 \rangle$ crystalline direction in order to minimize ion channeling. The system incorporates a mass spectrometer by which impurity ions are separated from that portion of the ion beam striking the samples.

Following implantation, the samples were cleaned again in the same way prior to encapsulation. The protective film deposited on the surface was either silicon nitride or silicon dioxide, depending on the tests being run. In either case, the samples are placed in a pyrolytic reactor where the requisite gases are introduced and react on the heated GaAs to produce a film. The sample is heated to 695°C in about five seconds by means of a carbon-strip heater. In the case of silicon nitride, the reacting gases are ammonia and silane (SiH_4), each diluted in nitrogen; in the case of silicon dioxide, they are silane and oxygen. In both cases, the system is first purged after evacuation with purified nitrogen gas taken from a tank of liquid nitrogen. Caps grown in this way were typically 1000 Å thick, requiring a reaction time of about 50 seconds.

The capped specimens were next annealed in flowing hydrogen gas for times and temperatures determined by the investigations. Temperatures ranged from 700°C to 900°C , and times from 3 minutes to 15 minutes. The caps were removed by etching in 48% hydrofluoric acid for 2-3 minutes (Si_3N_4) or for 30 seconds (SiO_2), and then recleaned with a water and methanol wash.

Electrical contacts were then made on the four corners of each sample using indium solder applied with the help of an ultrasonic soldering gun. In most cases, slight deviations from Ohmic behavior resulted. Contacts were therefore rendered Ohmic by annealing again at low temperature (300°C) for 5 minutes in an Argon flow.

Electrical Measurements

Hall effect and sheet resistivity measurements were made using the standard van der Pauw technique. In order to handle high resistivity materials, the system employed utilizes high impedance electrometers operated as unity-gain amplifiers. The output drives the shields on the leads between amplifier and sample, resulting in minimal leakage currents

and a low system time constant.

The van der Pauw technique is particularly useful since sample dimensions and contact separation do not enter into the calculations. All that is required are four contacts anywhere on the periphery of the sample. The method requires measurement of current and voltage corresponding to different pairs of contacts. These pairs must then be interchanged and the measurements averaged in order to correct for geometry. Measurements are also averaged in each case for forward and reverse sample currents and magnet currents in order to cancel stray galvanomagnetic and thermomagnetic effects which can give rise to spurious voltages superimposed on the Hall voltage to be measured.

From these measurements, the sheet-resistivity ρ_s (resistivity per unit thickness) and the sheet Hall coefficient R_{HS} (Hall coefficient per unit thickness) are determined directly. Hall mobility μ_H and sheet carrier concentration can then be calculated from the relations:

$$\mu_H = R_{HS} / \rho_s, \quad N_s = r / e R_{HS}$$

where r is the ratio of Hall mobility to conductivity mobility (usually taken as unity) and e is the electronic charge. N_s represents the sheet carrier concentration (ions/cm²).

Repeated measurements can be made on the surface of the sample after successive removal of measured layers by controlled etching. Thin layers can be stripped away uniformly and without damage using an etch solution of H₂SO₄:30% H₂O₂:H₂O in proportions of 1:1:50. At this concentration a typical etch rate is about 200 Å/minute as measured by a Sloan Dektak surface profile measuring system on a control sample placed alongside the measured sample. Contacts and leads were protected from the etchant by coating with black wax dissolved in trichloroethylene.

The carrier concentrations (per unit volume) and the mobilities in the i -th layer can then be obtained from the relations:

$$N_i = \frac{\Delta (1/\rho_s)_i}{e \mu_i d_i} \quad \text{and} \quad \mu_i = \frac{\Delta (R_{HS}/\rho_s^2)_i}{\Delta (1/\rho_s)_i}$$

where $\Delta (1/\rho_s)_i = \frac{1}{(\rho_s)_i} - \frac{1}{(\rho_s)_{i+1}}$

and $\Delta (R_{HS}/\rho_s^2)_i = \frac{(R_{HS})_i}{(\rho_s)_i^2} - \frac{(R_{HS})_{i+1}}{(\rho_s)_{i+1}^2}$

Here $(R_{HS})_i$ and $(\rho_s)_i$ are the measured sheet Hall coefficient and sheet resistivity after removal of the i -th layer of thickness d_i .

The magnetic field strength used was 7.6 Kgauss throughout. Sample currents ranged from milliamps to fractions of a microamp.

DISCUSSION AND RESULTS

I. Germanium-Implanted GaAs

Background and Rationale

Very little data exists on Germanium (Ge)-implanted GaAs. The only article since 1976 appears to be one by R.K. Surridge and B.J. Sealy of the University of Surrey (J. Phys. D: Appl. Phys 10 (1977)). This article reports n-type activity for 200 and 300 KeV Ge ions, annealed at 700°C for 15 minutes using an Aluminum cap. Doses ranged from 10^{13} to 10^{15} ions/cm². In particular, for a dose of 2×10^{14} ions/cm² and energy 200 KeV, these investigators found a sheet mobility of about 3600 cm²/V-sec, a sheet carrier concentration of 5×10^{12} /cm², and an activation efficiency of only 2.5%.

Germanium, like Silicon, is an amphoteric dopant in GaAs, with the possibility of producing both n or p type activity depending on implant and heat treatment conditions. It is desirable to determine the experimental factors critical to each behavior.

Penetration of Ge-implanted ions into GaAs should be relatively shallow. At an energy of 120 KeV, the peak concentration should occur at only 500 Å deep. Thus Ge implants may be useful in providing good, shallow n-type layers.

Ge implants also provide the possibility of making good ohmic contacts on n-type material. The most common method

at present is to evaporate a Au-Ge (88:12) alloy, followed by a Ni layer, and then the formation of a contact at around 450°C. The function of the Nickel is to reduce the cohesion of the Au-Ge alloy, preventing it from "balling up". The fast-diffusing Au drives the Germanium into the material as a donor. Ion implantation offers the alternate possibility of imbedding the Ge ions directly, hopefully making them substitutional at a substrate temperature significantly less than 450°C.

Results

The presence of a Ge-implanted layer in the samples used was confirmed by the method of glow discharge optical spectroscopy (G.D.O.S.), which looks at the luminescence of a glow discharge while sputtering away the surface. A Germanium line was seen appearing from about 400-500 Å deep during the sputtering.

Germanium-implanted samples (120 KeV) in six doses of 5×10^{12} , 1×10^{13} , 3×10^{13} , 1×10^{14} , 3×10^{14} , and 1×10^{15} ions/cm² were all encapsulated with Si₃N₄ and annealed for 15 minutes at 800°C. Results indicated that this anneal temperature was too low to achieve sufficient activation. In the samples where activation was measurable, activity was p-type with efficiencies of less than 8%.

Another set of samples was prepared in the same way but annealed at 900°C for 15 minutes. The two highest doses showed considerable cap failure during anneal, the cap stability improving inversely with the dose level. Electrical

measurements made on the four lowest doses showed p-type activity with good activation, ranging from 28% at $1 \times 10^{14}/\text{cm}^2$ to 90% at $5 \times 10^{12}/\text{cm}^2$. Mobilities ranged from 50 to 182 in these samples. These activations and mobilities showed so much improvement over results at 800°C anneal that they were suspect. An investigation of the substrate alone, processed through identical capping and annealing stages, was planned in order to determine whether conducting surface layers are produced on the semi-insulating GaAs alone.

This study was carried out on substrate samples from two different boules. Results showed p-type material with a wide variation in the value of average resistance measured directly across the corners of the sample surface, ranging from several kilohms to megohms. An effort to re-examine the operation and flow rates of the pyrolytic reactor was undertaken, since it was believed that defective capping was allowing changes in the substrate through out-diffusion of Arsenic, loss of Chromium, or some other effect. If the p-type substrate, as a result, has an apparent sheet concentration comparable to or greater than that of the ion-implanted layer, or if its resistivity is comparable to or less than that of the ion-implanted layer, then meaningful results descriptive of the ion layer itself cannot be found by electrical measurements. When the caps had improved, with more consistent resistances in the megohm range, the investigation proceeded.

It was felt that an intermediate anneal temperature around 850°C might produce sufficient activation of the ion

implants and at the same time prevent excessive deterioration of the substrate itself. The set of doses was annealed at 850°C and measured. Activation efficiencies were found to range from 7 to 40 %, and mobilities from 10 to 144 cm²/V-sec. The 3x10¹³/cm² dose sample, with 15% activation and mobility of 113 was selected for profiling. The profile of this sample however showed a concentration of carriers almost flat from the surface into the sample to a depth of around 1300 Å. If these measurements correspond to carriers due to activated Ge atoms, then they would indicate considerable diffusion of the implanted profile during the capping and 15 minute anneal stages.

A second Ge-implanted sample annealed at 900°C was also profiled, but showed similar behavior, reasonably flat at a concentration of about 3 x 10¹⁷/cm³, and then dropping off at a depth of about 1700 Å.

In all cases, the Ge-implanted layers showed p-type activity, unlike the results of Surridge and Sealy quoted earlier. At this point it is not possible to pinpoint the critical difference in preparation leading to opposite activities. More investigation of Germanium-implanted GaAs will be necessary before the relation between implant and anneal parameters and electrical properties can be described with confidence. At this stage, the data collected is sometimes inconsistent with previous data. However, the data described previously was found to be generally reproducible in samples tested later. Some of this data is summarized in the accompanying Table. Undoubtedly the variations in quality and

SUMMARY: HALL/RESISTIVITY MEASUREMENTS

15 June 1978 900°C Anneal		Type	ρ_s	R_{HS}	μ_H	N_s	Activation Efficiency
Dose:	5E12	P	1.2E4	1.4E6	112	4.5E12	90%
	1E13	P	5.1E3	9.2E5	182	6.8E12	68%
	3E13	P	3.0E3	4.8E5	161	1.3E13	43%
	1E14	P	4.6E3	2.2E5	49	2.8E13	28%
29 June 1978 850°C Anneal							
Dose:	5E12	P	2.2E4	3.1E6	144	2.0E12	40%
	1E13	P	7.5E4	1.7E6	22	3.8E12	37%
	3E13	P	1.3E4	1.4E6	113	4.3E12	14%
	1E14	P	2.3E4	4.9E5	21	1.3E13	13%
	3E14	P	3.6E4	3.2E5	9	2.0E13	6.5%
11 July 1978 900°C Anneal							
Dose:	3E13	P	4.5E3	6.9E5	154	9.0E12	30 %
	1E14	P	2.2E3	2.4E5	109	2.6E13	26 %
26 July 1978 900°C Anneal							
Dose:	1E13	P	4.5E4	6.5E5	14	9.6E12	96%
	3E13	P	2.6E4	6.0E5	23	1.0E13	35%
	1E14	P	2.6E3	2.6E5	101	2.4E13	24%
	3E14	P	7.7E3	1.4E6	182	4.4E12	1.5%

consistency of the caps, and in the substrate material itself, are chiefly responsible for discrepancies in measured electrical properties.

II. Magnesium-Implanted GaAs

Background and Rationale

In a study of Magnesium-implanted GaAs conducted by Dr. Y.K. Yeo of this laboratory, it was determined that Magnesium ions have large diffusion coefficients, producing rapid changes in their diffusion profile during heat treatment. A portion of this investigation directed to the determination of the magnitude of the diffusion coefficient for Magnesium was undertaken as part of this project and will be briefly reported here.

The basic theoretical framework which predicts the concentration distribution with depth (profile) of implanted ions is due to the unified range theory of Lindhard, Scharff, and Schiott (LSS theory). This theory produces for a given ion, substrate material, and ion dose & energy, values of the mean projected ion range R and a standard deviation in ion range R_s , which are fitted to a Gaussian distribution of the form:

$$n(x) = \frac{\text{Dose}}{\sqrt{2\pi} R_s} e^{-\frac{(x-R)^2}{2R_s^2}}$$

where n is ion density and x is distance from the surface.

Once implanted, ion diffusion occurs which alters the theoretical profile to a diffused profile. The latter, which depends on thermal treatment of the sample, can be shown to be describable analytically by replacing R_s in the above by the quantity $\sqrt{R_s^2 + 2Dt}$, where D is the diffusion coefficient (function of temperature) in cm^2/sec and t is the diffusion time. This solution to the diffusion equation assumes no surface condition affecting out-diffusion there.

Thus the measured concentration profile can be fitted to a Gaussian distribution by appropriate selection of diffusion coefficient D corresponding to anneal time t at a given anneal temperature. A study of diffusion profiles for a given temperature but different times then provides a method of determining diffusion coefficients for a particular implanted ion.

In the case of the Magnesium-implanted ions, Dr. Yeo had found that diffusion was so rapid that considerable change in the implanted profile already occurs during the short heat treatment involved in capping alone. With Si_3N_4 capping this treatment is at a temperature of 695°C for a time of about 45 seconds. Because of this preliminary diffusion, subsequent anneals up to temperatures of 700°C were found to make no appreciable alteration in the electrical behavior of the samples. The present study endeavored to use a SiO_2 cap which could be put down at a much lower temperature of 325°C in order to permit a diffusion rate study at temperatures lower than 700°C . Since Magnesium implanted GaAs is p-type, the SiO_2 cap would not present any difficulties, especially considering the low anneal temperatures planned. Diffusion behavior at

higher temperatures was already available from Dr. Yeo's measurements.

Results

This effort was only begun in the time available. Two doses of Mg-implanted GaAs were capped with about 1200 Å of SiO_2 and then annealed at three different anneal times at a temperature of 700°C .

Each sample was to be profiled and compared to results for the same dose and anneal temperature but a different anneal time in order to determine the diffusion coefficient. Since a single profile requires a series of Hall/resistivity measurements occupying a full day's time, only a few of these profiles could be made. The results of the profiles to date show good agreement with the previous measurements of Dr. Yeo.

In continuing this effort further, several future lines of investigation have been suggested by the work to date. A continuation of the diffusion study for Mg-implanted ions should include additional profiling with silicon nitride caps at other temperatures, especially at 700°C anneal. This will permit a better understanding of the effect of the capping material on diffusion profiles. In addition, the effect on the profiles of the Ohmic anneal itself should be systematically studied.